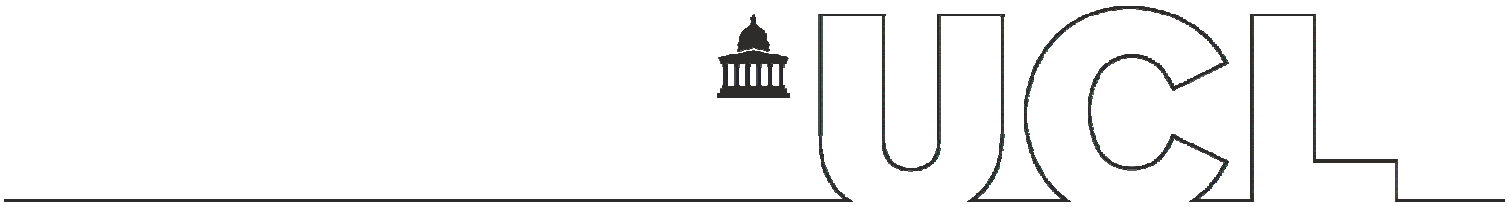


Dynamic models of brain imaging data and their Bayesian inversion

by

André Sousa Cardoso da Costa Marreiros

Wellcome Trust Centre for Neuroimaging
Institute of Neurology

A thesis submitted for the degree of Doctor of Philosophy
University College London
April, 2009

Primary supervisor: Professor Karl J Friston
Secondary supervisor: Dr. Stefan J Kiebel

Declaration

I, André Sousa Cardoso da Costa Marreiros, confirm that the work presented in this thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

"Knowledge is not a series of self-consistent theories that converge toward an ideal view; it is rather an ever increasing ocean of mutually incompatible (and perhaps even incommensurable) alternatives, each single theory, each fairy tale, each myth that is part of the collection forcing the others into greater articulation and all of them contributing, via this process of competition, to the development of our consciousness." - Paul Feyerabend

Abstract

This work is about understanding the dynamics of neuronal systems, in particular with respect to brain connectivity. It addresses complex neuronal systems by looking at neuronal interactions and their causal relations. These systems are characterized using a generic approach to dynamical system analysis of brain signals - dynamic causal modelling (DCM). DCM is a technique for inferring directed connectivity among brain regions, which distinguishes between a neuronal and an observation level. DCM is a natural extension of the convolution models used in the standard analysis of neuroimaging data. This thesis develops biologically constrained and plausible models, informed by anatomic and physiological principles. Within this framework, it uses mathematical formalisms of neural mass, mean-field and ensemble dynamic causal models as generative models for observed neuronal activity. These models allow for the evaluation of intrinsic neuronal connections and high-order statistics of neuronal states, using Bayesian estimation and inference. Critically it employs Bayesian model selection (BMS) to discover the best among several equally plausible models. In the first part of this thesis, a two-state DCM for functional magnetic resonance imaging (fMRI) is described, where each region can model selective changes in both extrinsic and intrinsic connectivity. The second part is concerned with how the sigmoid activation function of neural-mass models (NMM) can be understood in terms of the variance or dispersion of neuronal states. The third part presents a mean-field model (MFM) for neuronal dynamics as observed with magneto- and electroencephalographic data (M/EEG). In the final part, the MFM is used as a generative model in a DCM for M/EEG and compared to the NMM using Bayesian model selection.

Acknowledgments

First of all, I would like to thank to my principal and secondary supervisors, Karl Friston and Stefan Kiebel for their support and guidance over the past years. I owe immense gratitude to Karl for his inspiration, assistance and generosity. I thank him also for sharing his wide knowledge and insights. It has been exceptionally great to have him as my first supervisor. I also want to thank Stefan enormously. He gave me great and valuable support over these years, in many stages of this thesis and in many practical issues. Furthermore, I want to thank Jean Daunizeau for his always happy companionship, for the fun coffee breaks, for the integrative insights and for his help. I also want to thank Lee Harrison, who gave me kind support, cool hints, funny British humour and for always being so pleasant to work with.

Moreover, I want to thank to my examiners and to everyone who gave me feedback on my work. And also a special thank to Marcia Bennett for making many travel arrangements and for managing daily life at FIL in general easier. Not last, I want to state a special thank to all the fellows and PhD students at the methods group with whom I have build good friendships and spend many enjoyable moments.

I also want to thank to all my colleagues at the FIL for the friendly and scientifically great environment. I really feel that I have been lucky to be in such a nice, dynamic and vibrant place to work, referring especially to FIL but also to the ION.

I am also really grateful to my closest friends for being there and sharing life with me, without whom most things wouldn't make much sense. And last, but definitely first, I am in debt to my family for their always present support and unconditional love.

This work was funded by the *Portuguese Foundation for Science and Technology* (SFRH/BD/15858/2005) and the *Wellcome Trust*.

Table of Contents

Abbreviations

A1	primary auditory cortex
BF	Bayes factor
BMS	Bayesian model selection
BOLD	blood oxygenation level dependent
BS	brainstem
DCM	dynamic causal modelling
EEG	electroencephalography
EM	expectation-maximization
ERP	event-related potential
fMRI	functional magnetic resonance imaging
FN	FitzHugh-Nagumo
GLM	general linear model
HH	Hodgkin-Huxley
IFG	inferior frontal gyrus
MAR	multivariate autoregressive model
MEG	magnetoencephalography
MFM	mean-field model
MM	method of moments
MMN	mismatch negativity
MN	median nerve
MRI	magnetic resonance imaging
NMM	neural mass model
ODE	ordinary differential equation
PAS	paired associative stimulation
PET	positron emission tomography
PPI	psychophysiological interaction
SEM	structural equation modelling
SPC	superior parietal cortex
SPM	statistical parametric mapping
SDE	stochastic differential equation
SEP	sensory evoked potential

SSEP	somatosensory evoked potential
STG	superior temporal gyrus
SVD	singular value decomposition
TMS	transcranial magnetic stimulation

Figures and Tables

Outline and aims of this thesis

In the scientific study of the nervous system, one finds disciplines as diverse as cognitive and neuro-psychology, computer science, statistics, physics, philosophy, and medicine. The arrival of computers as tools for dealing with complex electrophysiological, molecular and image datasets in the 1970s has been followed by the increasing use of computer modeling and computer simulations of many brain functions.

The principal area of investigation in this work concerns the interface between imaging neuroscience and theoretical neurobiology. Mathematical techniques are developed to characterise brain organisation. This involves creating models of how the brain is wired and how it responds in different contexts. These models are used to interpret measured brain responses using brain imaging and electromagnetic brain signals.

Investigating the involvement of brain regions in various cognitive and perceptual tasks has become increasingly common in neuroimaging studies. Functional magnetic resonance imaging (fMRI) studies are especially popular, due to their non-invasive nature and high spatial resolution, likewise, electroencephalography (EEG) and magnetoencephalography (MEG) are popular, due to their non-invasive nature and high temporal resolution. Advances in data analysis and modelling make possible the use of these neuronal data to ask not only which brain regions are involved in these tasks, but also how they communicate with one another.

There is a broad consensus in neuroscience that mathematical system models are extremely helpful in neuroscience, for a mechanistic understanding of neural systems. Models of effective connectivity, i.e. the causal influences that system elements exert over another, are essential for studying the functional integration of neuronal populations and for understanding the mechanisms that underlie neuronal dynamics (Friston, 2002a; Horwitz et al., 1999). DCM is currently probably the most advanced and general framework for inferring processes and mechanisms at the neuronal level from measurements of functional neuroimaging data, including fMRI (Friston et al., 2003), EEG/MEG (David et al., 2006a) and local field potentials (Moran et al., 2007).

In contrast to other models of effective connectivity, DCM does not operate on the measured time-series directly. Instead, it combines a model of the hidden neuronal dynamics with a forward model (or generative model) that translates neuronal states into predicted measurements of how observed data were caused.

DCM can be used to infer whether neuronal functional coupling is modulated by experimental manipulations, like task demands, stimulus properties, learning, attention, drugs, etc… The coupling amongst neuronal populations changes as a function of processing demands (McIntosh, 2000; Stephan, 2004). We hope that in the next years, the generic framework of DCM and related developments, will contribute to a more mechanistic understanding of brain function; to help understand the mechanisms of drugs and to develop models that can serve as diagnostic tools for diseases linked to abnormalities of connectivity and synaptic plasticity, e.g. Schizophrenia and Parkinson. Another possibility is to explore its utility as a diagnostic tool. The obvious extension of DCM is in terms of its neurophysiological plausibility. Thus, the aim of this thesis is to endow DCMs with a greater biological realism informed by anatomical and physiological constrains. The first part of the work described in this thesis focuses on excitatory-inhibitory DCM models for fMRI time series. In the second part, specific questions are formulated and addressed concerning the role of variance in DCM for ERPs. Bayesian model selection (BMS) is the key for selecting among competing models and hypotheses.

Research within this work has been mostly concerned with the dynamical system aspect of the neuronal interactions among brain areas, within the DCM framework. This is done by using previously developed procedures in Bayesian estimation and inference for dynamic causal models and adapting those methods to the new models developed here. The ‘creative’ work in this thesis is the development of novel generative models and mechanisms to describe key aspects of functional neuroimaging data. This thesis comprises seven chapters and a number of appendices. The first two chapters introduce the domain of neural imaging models and the background framework of DCM. Chapters 3-6 contain the main results, which are concluded by overall discussion on Chapter 7. This thesis is structured as follows:

Chapter 1 introduces functional neuroimaging and neuronal structure–function relationships. It goes through the two main principles of functional brain organisation: functional segregation and integration. It shows how functional integration is usually analysed in terms of functional or effective connectivity models. While functional connectivity describes statistical dependencies between data, effective connectivity rests on a mechanistic model of the causal effects that generated the data.

Chapter 2 presents a short review on Dynamical Causal Modelling as a technique for determining the effective connectivity in neural systems. It considers neuronal causal models, then introduces the bilinear models for fMRI time series and neural mass model (NMM) for ERPs. As measured with EEG/MEG. Finally, it presents the basis of Bayesian model selection.

Chapter 3 endows dynamic causal models (DCM) for fMRI time series with a greater biological realism. It presents the theory, methods and implementation of an extension of dynamic causal modelling to include, region specific excitatory and inhibitory neural populations in networks of coupled neural masses. Critically, the extension allows us to place positivity constraints on the connectivity such that the model conforms to a more realistic organisation of cortical hierarchies, whose extrinsic connections are excitatory (glutamatergic). Consequently, we can model changes in both extrinsic and intrinsic connectivity.

Chapter 4 concerns the effect of dispersion (variance) of neuronal states on the cortical responses to sensory inputs. It provides a link between the sigmoid activation function and the variance of neuronal membrane depolarization, through a cumulative density function within a population. This Chapter lays the ground-work of the extension in the following two Chapters whereby the variance itself is a time-dependent variable and hence dynamically coupled to the mean. This provides a crucial link between neural mass models and more general neural density models.

Chapter 5 elucidates and generalizes the connections between moment equations (mean, variance, *etc*) and the full ensemble description for neural states. It develops a generic mean-field treatment of neuronal dynamics, which is based on a Laplace

approximation to the ensemble density and is formulated in terms of equations of motion for the sufficient statistics (i.e., the mean or mode and the variance or dispersion) of the ensemble density. This approach reduces to a neural-mass model when the second-order statistics (variance) of neuronal states are ignored. The interesting key behaviour is the coupling between the mean and variance of the ensemble, which is lost in the neural-mass approximations. Results of the mean-field method are compared numerically to the neural-mass method, in which only the mean is included.

Chapter 6 describes and evaluates a DCM based on density-dynamics instead of neural-mass models. It uses the Laplace and neural mass approximations as generative models of electrophysiological responses to sensory input. The role of higher moments is assessed empirically in a Bayesian model selection framework and the evidence for a role of the variance in shaping population dynamics is considered.

Chapter 7 provides a general discussion and the conclusions of this work and indicates directions for exciting future research.

The references are preceded by appendices containing relevant scientific material (**Appendices A**, **F**) and technical details (**B**, **C**, **D**).

CHAPTER 1

INTRODUCTION: MODELS IN NEUROIMAGING

The aim of this Chapter is to introduce the key models used in imaging neuroscience and to see how they relate to each other. The Chapter begins by introducing functional specialization and integration concepts, and anatomical models of functional brain architectures, which motivate some of the fundaments in neuroimaging. It briefly goes through some basic statistical models used for making classical and Bayesian inferences about *where* neuronal responses are expressed. By incorporating biophysical constraints, these basic models can be finessed and, in a dynamic setting, rendered causal. This allows us to infer *how* interactions among brain regions are mediated. Brain responses models are briefly reviewed, starting with the general linear model (GLM) of functional magnetic resonance imaging (fMRI). This model is successively refined until we arrive at effective connectivity models, like DCM, which will be the focus of next Chapter. Most of this material is based on work from my supervisor, Karl Friston.

1.1 Functional Imaging

Functional neuronal imaging studies human brain function based on analysis of data acquired using brain imaging modalities such as Electroencephalography (EEG), Magnetoencephalography (MEG), functional Magnetic Resonance Imaging (fMRI), Positron Emission Tomography (PET) or Optical Imaging. The aim is to understand the brain mechanics at multiple spatial and temporal scales, in terms of its physiology, functional architecture and dynamics. The framework for these studies includes classical techniques from neuroanatomy, neurophysiology, and the cognitive neurosciences, as well as perspectives from computational and theoretical neuroscience and physics.

Modern functional imaging has two main advantages over the multi/single-unit recordings used to study the electrophysiology of neurons. The first is that it is

generally non-invasive, and is therefore applicable routinely in humans. This allows for the study of unique human attributes such as language. The second is that it can acquire simultaneous activity from the whole brain. Compared to single or multiple neuron measurements, these large-scale brain observations at a systems level provide a different yet complementary perspective on neural coding (see *e.g.,* functional integration, below). A disadvantage, however, is that functional imaging provides only an indirect measure of the quantities of primary interest to neuroscientists *e.g*., firing rates and membrane potentials. There is current research which aim at bridging this gap using a combination of experimental and mathematical modelling approaches (Bertrand and Tallon-Baudry, 2000; Foster et al., 2008; Shulman et al., 2002).

1.1.1 Functional Segregation, Specialization and Integration

From a historical perspective, the distinction between functional specialisation and functional integration relates to the dialectic between *localisationism* and *connectionism*, which dominated thinking about brain function in the nineteenth century. Since the formulation of phrenology by Franz Gall (around 1800), who postulated fixed one-to-one relations between particular parts of the brain and specific mental attributes, the identification of a particular brain region with a specific function has become a central theme in neuroscience. During the following decades, lesion and electrical stimulation paradigms were developed to test whether functions could indeed be localised in animal models.

In 1881, Friedrich Goltz, although accepting the results of electrical stimulation in dog and monkey cortex, held a unitary view of brain function. He considered that the excitation method was inconclusive, in that the movements elicited might have originated in related pathways, or current could have spread to distant centres (Phillips et al., 1984). In short, the excitation method could not be used to infer functional localisation because localisationism discounted interactions, or functional integration among different brain areas. Though, only some years later, observations on patients with brain lesions (Absher, 1993) led to the concept of *disconnection syndromes* and the refutation of localisationism as a complete or sufficient explanation of cortical organisation.

Functional localisation implies that a function can be localised in a cortical area, whereas specialisation suggests that a cortical area is specialised for some aspects of perceptual or motor processing, and that this specialisation is anatomically *segregated* within the cortex. The cortical infrastructure supporting a single function may then involve many specialised areas whose union is mediated by the functional integration among them. In this view, functional specialisation is only meaningful in the context of functional integration and *vice versa*.

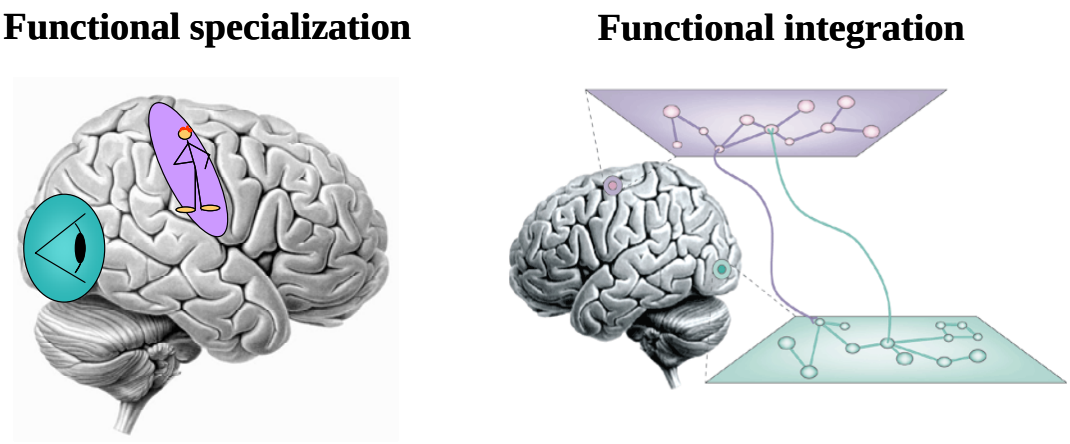


Figure 1.1: Schematic of the principles of brain organization. Functional specialization (left) refers to the existence of specialized neurons and brain areas, organized into distinct neuronal populations and grouped together to form segregated cortical areas. Functional integration (right) refers to interactions between distant neuronal units or networks from different parts of the brain. The interplay of segregation and integration in brain networks generates patterns of high complexity, which enable the emergence of coherent cognitive and behavioural states. Adapted from (Varela et al., 2001).

1.1.2 Models of functional specialization of regionally specific responses

Functional mapping studies are usually analysed with some form of statistical parametric mapping to test hypotheses about regionally specific effects (Friston et al., 1991). Statistical parametric mapping (SPM) is a voxel-based approach, employing classical statistics and topological inference, to make comments about regionally

specific responses to experimental factors. PET or fMRI data are first spatially processed so that they conform to a known anatomical space, in which responses are characterized statistically; typically using the General Linear Model (GLM) (Friston, 1995a).

For fMRI data the GLM embodies a convolution model of the haemodynamic response (Boynton et al., 1996; Friston, 1994). This accounts for the fact that BOLD signals are a delayed and dispersed version of the neuronal response. GLMs are fitted at each voxel and inferences are made about which parts of the brain are active, in a statistical sense. To accommodate the spatial nature of the imaging data (and account for the multiple statistical comparisons made) SPM techniques make use of Random Field Theory (RFT) (Worsley et al., 1996), and/or other statistical procedures, e.g., False Discovery Rate (Genovese et al., 2002).

There is also a Bayesian alternative to classical inference with SPMs, which looks at conditional inferences about an effect, given the data, as opposed to classical inferences about the data, given the effect is zero. Bayesian inferences about effects that are continuous in space use Posterior Probability Maps (PPMs) (Friston et al., 2002). Although not as widely used as SPMs, PPMs are potentially very useful, not least because they do not have to contend with the multiple-comparisons problem induced by classical inference (Berry and Hochberg, 1999).

Alternatively, MEG or EEG data can be analyzed to furnish a crude spatial mapping of brain function. Functions can, however, be more accurately localized using source reconstruction methods (Baillet et al., 2001). This entails specifying a forward model describing how a current source in the brain propagates to become an MEG or EEG measurement, using Maxwell's equations. These models are then inverted using statistical inference. Data from sensory systems are often analyzed using an averaging procedure. The data immediately following a sensory event, *e.g.,* hearing an auditory tone, are averaged over multiple events to produce an Event Related Potential (ERP). Components of the ERP can then be localized to different parts of the brain. Other cognitive components, however, are not easily isolated using this ERP approach. For these, a time-frequency characterization may be more appropriate (Tallon-Baudry and

Bertrand, 1999). See also (Makeig et al., 2002) for a critique of the averaging procedure.

The SPM approach can also be used with structural data, acquired using structural MRI, to find brain regions containing a higher gray matter density. This is known as Voxel-Based Morphometry (VBM) (Ashburner and Friston, 2000) and has been used, for example, to show that the posterior hippocampus, engaged by spatial navigation, is enlarged in taxi drivers (Maguire et al., 2000).

1.1.3 Anatomical connectivity

The cortex is a complex system, characterized by its dynamics and architecture, which underlie many functions such as action, perception, learning, language, and cognition. Anatomical or structural connectivity refers to the brain network design of physical connections linking sets of neuronal elements, and its associated structural biophysical attributes such as synaptic strength or effectiveness.

Neural connectivity patterns have long attracted the attention of neuroanatomists (Brodmann, 1909; Cajal, 1909; Swanson, 2003) and play crucial roles in determining the (functional) properties of neurons and neuronal systems. One key aspect of the complexity of the nervous system is its intricate morphology, especially the multi-interconnectivity of its neuronal processing elements. The anatomical connections are relatively stable for short time scales, such as seconds to minutes (Linden et al., 2003; Todd and Marois, 2004). Structural connectivity patterns for longer time scales (hours to days) are possible to be subject to significant morphological change and plasticity (Draganski et al., 2006; Trachtenberg et al., 2002).

Invasive tracing studies are capable of collectively demonstrating direct axonal connections. By contrast, diffusion weighted imaging techniques, such as DTI, are useful as whole brain *in vivo* markers of fibre tracts. For example, DTI has been used to identify three regions of human parietal cortex based on their connectivity patterns with other brain areas (Rushworth et al., 2006). Moreover, structural imaging can also be used clinically. The best established application is the use of MRI for pre-surgical

mapping to localize tissue within or near regions intended for neurosurgical resection (Matthews et al., 2006).

One could ask the rather provocative question: why does one need to know any anatomy? Would it be acceptable to simply infer the presence of an anatomical connection from the functional characteristics of a system? Actually, some knowledge of anatomy is important to define the "connectivity space", thereby providing a plausible biological framework for theories and inferences about neural interactions when analysing functional neuroimaging data and developing computer simulations. Brain connectivity can be described at different levels or scales. At a microscale, which includes individual synaptic connections that link individual neurons, at a mesoscale where networks connect neuronal populations, at a macroscale where brain regions are linked by fibre pathways. Each level of description relates to specific neuroscience data, from single-unit recordings, through local field potentials to functional magnetic resonance imaging (fMRI), electroencephalogram (EEG), and magnetoencephalogram (MEG). It is likely that anatomical variability is one of the main sources for functional variability, expressed in neural dynamics and behavioural performance.

Another question which could be raised is: what measurements of anatomical connectivity are most useful to the study of how the brain works? Knowing if there are direct connections between two neurons or cortical areas is clearly important, but a complete description of the connections includes information such as the receptor subtypes at synapses (e.g. AMPA *vs.* NMDA), the ratio of inhibitory to excitatory interneurons, the number of connections and their physiological impact (modulatory top-down *versus* driving bottom-up inputs). At this point the boundary between anatomy and function becomes blurred. For example, it is almost impossible to distinguish among macaque V2-5 cortical areas using anatomical criteria alone, so these regions might be better classified if they were divided according to their responses to physiological stimuli rather than their morphology (Kisvárday, 1996).

1.2 Models of Functional Integration

Imaging neuroscience has firmly established functional specialisation as a principle of brain organisation in man. However, the integration of specialised areas has proven more difficult to assess. In 'functional integration' models are used to describe how different brain areas interact. A classic example is the use of models to find increased connectivity between dorsal and ventral visual streams after subjects learn object-place associations (Lerner et al., 2002; Ungerleider and Haxby, 1994). In fact, a wide range of statistical techniques are used to measure inter-regional connectivity. Both unsupervised (e.g., Independent Component Analysis, ICA; Brown et al., 2001) and supervised techniques (e.g., support vector machine, SVM; (Mourao-Miranda et al., 2005)) are used. Other models seek to directly measure "causal" connectivity based on static, statistical constraints (e.g., Structural Equation Modelling, SEM; (McIntosh and Gonzalez-Lima, 1994)) or dynamic, through more bio-physically motivated assumptions (e.g., Dynamic Causal Modelling, DCM; (Friston et al., 2003)). A challenge for functional integration models is to bridge the gap between the large-scale, statistical models of the whole brain, and the small number of highly constrained spatial regions needed to be able to apply SEM and/or DCM.

Experimentally, one could also look at the combination of transcranial magnetic stimulation (TMS) with Neuroimaging, which allows the use of localized perturbations of brain networks while they are engaged in the performance of specific tasks (Massimini et al., 2005). The theory of directed graphs can also be applied to analysis of structural (fibre pathways), functional (correlations) and effective (information flow) brain connectivity at all levels (e.g., (Brandes, 2005; Wen and Chklovskii, 2005)).

1.2.1 Functional Connectivity

Characterising brain activity in terms of functional specialisation does not reveal anything about how different brain regions communicate with each other. Functional connectivity, in contrast, is defined as statistical dependencies or correlations *among*

remote neurophysiological events. Statistical dependence may be estimated by measuring correlation or covariance, spectral coherence or phase-locking.

(Friston et al., 1993) introduced a voxel-based principal component analysis (PCA) of neuroimaging time-series to characterise distributed brain systems implicated in sensorimotor, perceptual or cognitive processes. These distributed systems are identified with principal components or *eigenimages* that correspond to spatial modes of coherent brain activity. This approach represents one of the simplest multivariate characterisations of functional neuroimaging time-series and falls into the class of exploratory analyses. Principal component or eigenimage analysis generally uses singular value decomposition (SVD) to identify a set of orthogonal spatial modes that capture the greatest amount of variance expressed over time. As such the ensuing modes embody the most prominent aspects of the variance-covariance structure of a given time-series. Noting that covariance among brain regions is equivalent to functional connectivity renders eigenimage analysis particularly interesting because it was among the first ways of addressing functional integration (*i.e.* connectivity) with neuroimaging data. Subsequently, eigenimage analysis has been elaborated in a number of ways. Notable among these is canonical variate analysis (CVA) and multidimensional scaling (Friston et al., 1996a; Friston et al., 1996b). Canonical variate analysis was introduced in the context of MANCOVA (multiple analysis of covariance) and uses the generalised eigenvector solution to maximise the variance that can be explained by some explanatory variables relative to error. CVA can be thought of as an extension of eigenimage analysis that refers explicitly to some explanatory variables and allows for statistical inference.

In fMRI, eigenimage analysis (*e.g.* (Sychra et al., 1994)) is generally used as an exploratory device to characterise coherent brain activity. These variance components may, or may not be, related to experimental design. For example, endogenous coherent dynamics have been observed in the motor system at very low frequencies (Biswal et al., 1995). Despite its exploratory power, eigenimage analysis is limited for two reasons. Firstly, it offers only a linear decomposition of any set of neurophysiological measurements and second, the particular set of eigenimages or spatial modes obtained is determined by constraints that are biologically implausible. These aspects of PCA confer inherent limitations on the interpretability and

usefulness of eigenimage analysis of biological time-series and have motivated the exploration of nonlinear PCA and neural network approaches.

There are two other important approaches. The first is independent component analysis (ICA). ICA uses entropy maximisation to find, using iterative schemes, spatial modes or their dynamics that are approximately *independent*. This is a stronger requirement than *orthogonality* in PCA and involves removing high-order correlations among the modes (or dynamics). It was initially introduced as *spatial* ICA (McKeown et al., 1998) in which the independence constraint was applied to the modes (with no constraints on their temporal expression). More recent approaches use, by analogy with magneto- and electrophysiological time-series analysis, *temporal* ICA where the dynamics are enforced to be independent. This requires an initial dimension reduction (usually using conventional eigenimage analysis). Finally, there has been an interest in cluster analysis (Baumgartner et al., 1997). Conceptually, this can be related to eigenimage analysis through multidimensional scaling and principal co-ordinate analysis.

Demonstrating statistical dependencies among regional brain responses or endogenous activity (*i.e.,* demonstrating functional connectivity) does not tell one much about how the brain works. An alterative approach is to use multivariate observation models of regional responses; which are now being used more and more frequently. Multivariate models map from the causes of brain responses (encoding models; $g(\theta):X \rightarrow Y$) or from brain activity to its consequences (decoding models; $g(\theta):X \rightarrow Y$), (Friston et al., 2008). Although to ask specific questions about how brain responses are caused, one needs explicit models of integration or more precisely, effective connectivity.

1.2.2 Effective Connectivity

Effective connectivity may be viewed as the union of structural and functional connectivity, as it describes networks of directional effects of one neural element over another. In principle, causal effects can be inferred through systematic perturbations

of the system, or, since causes must precede effects in time, through time series analysis. Effective connectivity refers explicitly to *the influence that one neural system exerts over another*, either at a synaptic (*i.e.* synaptic efficacy) or population level. It has been proposed that "the [electrophysiological] notion of effective connectivity should be understood as the experiment- and time-dependent, simplest possible circuit diagram that would replicate the observed timing relationships between the recorded neurons" (Aertsen and Preissl, 1991).

Various techniques for extracting effective connectivity have been pursued, including regression models (Friston, 1993, 1995b; McIntosh et al., 1994), convolution models (Friston, 2002b; Friston and Büchel, 2000) and state-space models (Büchel and Friston, 1998). Regression techniques, underlying e.g. the analysis of psychophysiological interactions (PPIs, see *Appendix A*), are useful because they are easy to fit and can test for the modulatory interactions of interest (Friston et al., 1997).

However, simple regression-based techniques exclude temporal information, i.e. the history of an input or physiological variable. This is important as interactions within the brain, whether over short or long distances, take time and are not instantaneous. Structural equation modelling (SEM, see *Appendix A*), as used by the neuroimaging community (Büchel and Friston, 1997; McIntosh and Gonzalez-Lima, 1994) has similar problems. These static models discount temporal information. Consequently, time-permuted data produce the same path coefficients as the original data.

Models that use the order in which data are produced are more natural candidates for neuronal dynamics. Models that can address the temporal aspect of causality include convolution models, such as the Volterra approach, which model temporal effects in terms of an idealized response characterized by kernels or impulse response functions (Friston et al., 2000). A criticism of the Volterra approach is that it treats the system as a black box, meaning that it has no model of the internal mechanisms that may generate data.

State-space models account for correlations within the data by invoking state variables whose dynamics generates data. For example, dynamic SEM models which can model temporal information (Cudeck, 2002). Recursive algorithms, such as the Kalman filter, can be used to estimate states through time, given the data (Büchel and Friston,

1998). Multivariate autoregressive models (MAR), which focus on the causal dependence of the present on the past time series was first implemented for fMRI by (Harrison et al., 2003). A complementary MAR approach, based on the idea of 'Granger causality' (Granger, 1969), was proposed by (Goebel et al., 2003). In this framework, given two time-series y_1 and y_2, y_1 is considered to be caused by y_2 if its dynamics can be predicted better using past values from y_1 and y_2 as opposed to using past values of y_1 alone. Finally, there is dynamic causal modelling (DCM) which was first introduced as a technique for determining effective connectivity in neural systems of interest on the basis of measured fMRI data (Friston et al., 2003). DCM is the topic of this thesis and will be introduced in detail in the next chapter.

1.3 Conclusion

In this Chapter we have reviewed some key models that underpin image analysis and have touched briefly on ways of assessing specialization and integration in the brain. Functional specialization assumes that local computations are used in certain aspects of information processing. Functional integration can be characterized in two ways, namely in terms of functional connectivity and effective connectivity. While functional connectivity describes statistical dependencies between data, effective connectivity rests on a mechanistic model of the causal effects that generated the data.

CHAPTER 2

THEORETICAL BACKGROUND: DYNAMIC CAUSAL MODELLING

The previous Chapter introduced models in neuroimaging, focusing on two interrelated concepts; functional specialization and functional integration, which have been guiding neuroimaging applications over the last few decades. This Chapter focuses exclusively on a recently established technique for determining the effective connectivity in neural systems of interest: Dynamic causal modelling (DCM). DCM is a general framework for inferring processes and mechanisms at the neuronal level from measurements of brain activity with different techniques, including fMRI (Friston et al., 2003), EEG/MEG (David et al., 2006a) and frequency spectra based on local field potentials (Moran et al., 2007). Here we review the conceptual and mathematical basis of DCM and Bayesian model selection (BMS; (Penny et al., 2004a; Stephan et al., 2009)). Since the original description of DCM (Friston et al., 2003), a number of methodological developments have improved and extended DCM for fMRI, e.g. precise sampling from predicted responses (Kiebel et al., 2007b), additional states at the neuronal level (Marreiros et al., 2008b), a refined hemodynamic model (Stephan et al., 2007c) and a nonlinear neuronal model (Stephan et al., 2008). DCM for EEG/MEG has also seen some extensions since its origins (David et al., 2006a), DCM for induced responses (Chen et al., 2008), DCM for neural-mass and mean-field models (Marreiros et al., 2009), DCM for spectral responses (Moran et al., 2009) and a nonlinear stochastic DCM (Daunizeau, 2009).

2.1 General causal models of neuronal interactions

Effective connectivity requires a causal model of the interactions between the elements of a neural system of interest. DCM is a technique for determining effective connectivity in neural systems of interest on the basis of measured fMRI and EEG/MEG, which will be introduced here and in the next sections. The mathematical

framework of DCM comprises deterministic differential equations with time-invariant parameters. The underlying concept is quite general: a *system* is defined by a set of elements with n time-variant properties that interact with each other. Each time-variant property x_i ($1 \leq i \leq n$) is called a *state variable*, and the n vector $x(t)$ of all state variables in the system is called the *state vector* (or simply *state*) of the system at time t:

$$x(t) = \begin{bmatrix} x_1(t) \\ \vdots \\ x_n(t) \end{bmatrix}. \tag{2.1}$$

Taking an ensemble of interacting neurons as an example, the system elements would correspond to the individual neurons, each of which is represented by one or several state variables. These state variables could refer to various neurophysiological properties, e.g. postsynaptic potentials, status of ion channels, etc. Critically, the state variables interact with each other, i.e. the evolution of each state variable depends on at least one other state variable. For example, the postsynaptic membrane potential depends on which and how many ion channels are open; vice versa, the probability of voltage-dependent ion channels opening depends on the membrane potential. Such mutual functional dependencies between the state variables of the system can be expressed quite naturally by a set of ordinary differential equations (ODEs) that operate on the state vector:

$$\frac{dx}{dt} = \begin{bmatrix} f_1(x_1, \ldots, x_n) \\ \vdots \\ f_n(x_1, \ldots, x_n) \end{bmatrix} = F(x). \tag{2.2}$$

However, this description is not yet sufficient. First of all, the specific form of the dependencies f_i needs to be specified, i.e. the nature of the causal relations between state variables. This requires a set of parameters θ which determine the form and strength of influences between state variables. In neural systems, these parameters usually correspond to time constants or synaptic strengths of the connections between the system elements. The Boolean nature of θ, i.e. the pattern of absent and present

connections, and the mathematical form of the dependencies f_i represent the *structure* of the system. Second, for non-autonomous systems (i.e. systems that exchange matter, energy or information with their environment) we need to consider the inputs into the system, e.g. sensory information entering the brain. We represent the set of all m known inputs by the m-vector function $u(t)$. Extending Eq. 2.2 accordingly leads to a general state equation for non-autonomous deterministic systems:

$$\frac{\partial x}{\partial t} = F\left(x, u, \theta\right). \tag{2.3}$$

A model whose form follows this general state equation provides a causal description of how system dynamics result from system structure, because it describes (i) when and where external inputs enter the system; and (ii) how the state changes induced by these inputs evolve in time depending on the system's structure. Given a particular temporal sequence of inputs $u(t)$ and an initial state $x(0)$, one obtains a complete description of how the dynamics of the system (i.e. the trajectory of its state vector in time) results from its structure by integration of Eq. 2.4:

$$x\left(\tau\right) = x\left(0\right) + \int_0^\tau F\left(x, u, \theta\right) dt\,. \tag{2.4}$$

Equation 2.3 therefore provides a general form for models of effective connectivity in neural systems. As described elsewhere (Friston et al., 2003; Stephan, 2004), all established models of effective connectivity, including regression-like models (Harrison et al., 2003; McIntosh and Gonzalez-Lima, 1994), can be related to this general equation. The next section shows how DCM models neural population dynamics using a bilinear implementation of this general form. This is combined with a forward model that translates neural activity into a measured signal.

Before we proceed, it is worth pointing out that we have made two main assumptions in this section to simplify the exposition of the general state equation. First, it is assumed that all processes in the system are deterministic and occur instantaneously. Whether or not this assumption is valid depends on the particular system of interest. If necessary, random components (noise) and delays could be accounted for by using

stochastic differential equations (SDEs) and delay differential equations, respectively. An example of the latter is found in DCM for evoked responses (see below). Second, we assume that we know the inputs that enter the system. This is a tenable assumption in neuroimaging because the inputs are experimentally controlled variables, e.g. changes in stimuli or instructions. It may also be helpful to point out that using time-invariant dependencies f_i and parameters θ does not exclude modelling time-dependent changes of the network behaviour. Although the mathematical form of f_i *per se* is static, the use of time-varying inputs u allows for dynamic changes in what components of f_i are 'activated'. For example, input functions that can only take values of one or zero and that are multiplied with the different terms of a polynomial function can be used to induce time-dependent changes from nonlinear to linear behaviour (e.g. by "switching off" all higher order terms in the polynomial) or vice versa. Also, there is no principled distinction between states and time-invariant parameters. Therefore, estimating time-varying parameters can be treated as a state estimation problem.

2.2 Dynamic causal modelling with bilinear models

This section is about modelling interactions among neuronal populations, at a cortical level, using neuroimaging time-series and dynamic causal models that are informed by the biophysics of the system studied. The aim of DCM is to estimate, and make inferences about, the coupling among brain areas and how that coupling is influenced by experimental changes (*e.g.* time or cognitive set). The basic idea is to construct a reasonably realistic neuronal model of interacting cortical regions or nodes. This model is then supplemented with a forward model of how neuronal or synaptic activity translates into a measured response. This enables the parameters of the neuronal model (*i.e.* effective connectivity) to be estimated from observed data.

Intuitively, this approach regards an experiment as a designed perturbation of neuronal dynamics that are promulgated and distributed throughout a system of coupled anatomical nodes to change region-specific neuronal activity. These changes engender, through a measurement-specific forward model, responses that are used to identify the architecture and time constants of the system at a neuronal level. This

represents a departure from conventional approaches (*e.g.*, structural equation modelling and auto-regression models; (Büchel and Friston, 1997; Harrison et al., 2003; McIntosh and Gonzalez-Lima, 1994)), in which one assumes the observed responses are driven by endogenous or intrinsic noise (*i.e.* innovations). In contrast, dynamic causal models assume the responses are driven by designed changes in inputs. An important conceptual aspect of dynamic causal models pertains to how the experimental inputs enter the model and cause neuronal responses. Experimental variables can elicit responses in one of two ways. First, they can elicit responses through direct influences on specific anatomical nodes. This would be appropriate, for example, in modelling sensory evoked responses in early visual cortices. The second class of input exerts its effect vicariously, through a modulation of the coupling among nodes. These sorts of experimental variables would normally be more enduring; for example attention to a particular attribute or the maintenance of some perceptual set. These distinctions are seen most clearly in relation to particular forms of causal models used for estimation, for example the bilinear approximation

$$\begin{aligned} \dot{x} &= f(x,u,\theta) \\ &= Ax + uBx + Cu \\ y &= h(x,u,\theta) + \varepsilon \end{aligned} \tag{2.5}$$

$$A = \frac{\partial f}{\partial x} \qquad B = \frac{\partial^2 f}{\partial x \partial u} \qquad C = \frac{\partial f}{\partial u}$$

where $\dot{x} = \partial x / \partial t$. This is an approximation to any model of how changes in neuronal activity in one region x_i are caused by activity in the other regions. Here the output function $h(x)$ embodies a haemodynamic model, linking neuronal activity to BOLD, for each region (see Figure 2.2). θ are the quantities that parameterize the state and observer equations (A, B, C). The matrix A represents the coupling among the regions in the absence of input $u(t)$. This can be thought of as the latent coupling in the absence of experimental perturbations. The matrix B is effectively the change in latent coupling induced by the input. It encodes the input-sensitive changes in A or, equivalently, the modulation of coupling by experimental manipulations. Because B is a second-order derivative it is referred to as *bilinear*. Finally, the matrix C

embodies the extrinsic influences of inputs on neuronal activity. The parameters $\theta = A, B, C$ are the connectivity or coupling matrices that we wish to identify and define the functional architecture and interactions among brain regions at a neuronal level. Figure 2.1 summarises this bilinear state equation and shows the model in graphical form.

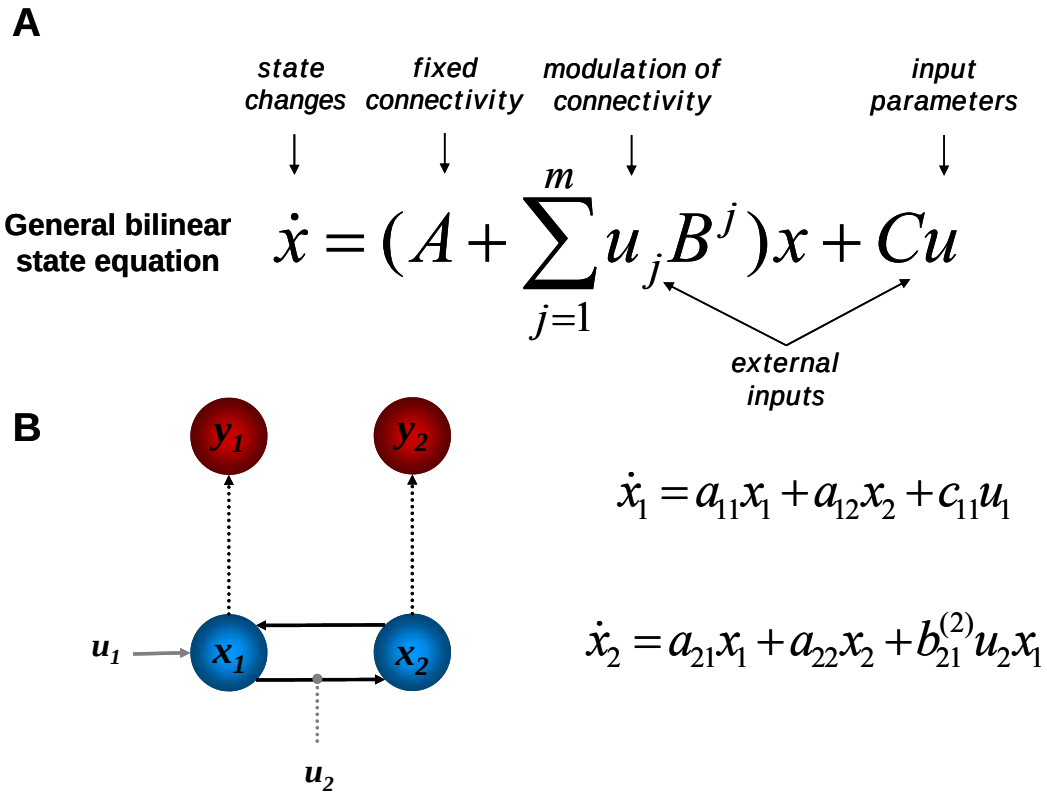


Figure 2.1: (A) The bilinear state equation of DCM for fMRI. (B) An example of a DCM describing the dynamics in a hierarchical system of visual areas. This system consists of two areas represented by a single state variable (x_1, x_2). Black arrows represent connections, grey arrows represent external inputs into the system and thin dotted arrows indicate the transformation from neural states (blue colour) into haemodynamic observations (red colour) (see figure 2.2 for the haemodynamic forward model). The state equation system for this particular model is shown on the right. Adapted from (Stephan et al., 2007a).

DCM for fMRI combines this model of neural dynamics with an experimentally validated haemodynamic model that describes the transformation of neuronal activity

into a BOLD response. This so-called "Balloon model" was initially formulated by (Buxton et al., 1998) and later extended by (Friston et al., 2000). Briefly, it consists of a set of differential equations that describe the relations between four haemodynamic state variables, using five parameters $\theta^{(h)}$. More specifically, changes in neural activity elicit a vasodilatory signal that leads to increases in blood flow and subsequently to changes in blood volume and deoxyheamoglobin content. The predicted BOLD signal is a non-linear function of blood volume and deoxyhaemoglobin content. This haemodynamic model is summarised in Figure 2.2 (Friston et al., 2000).

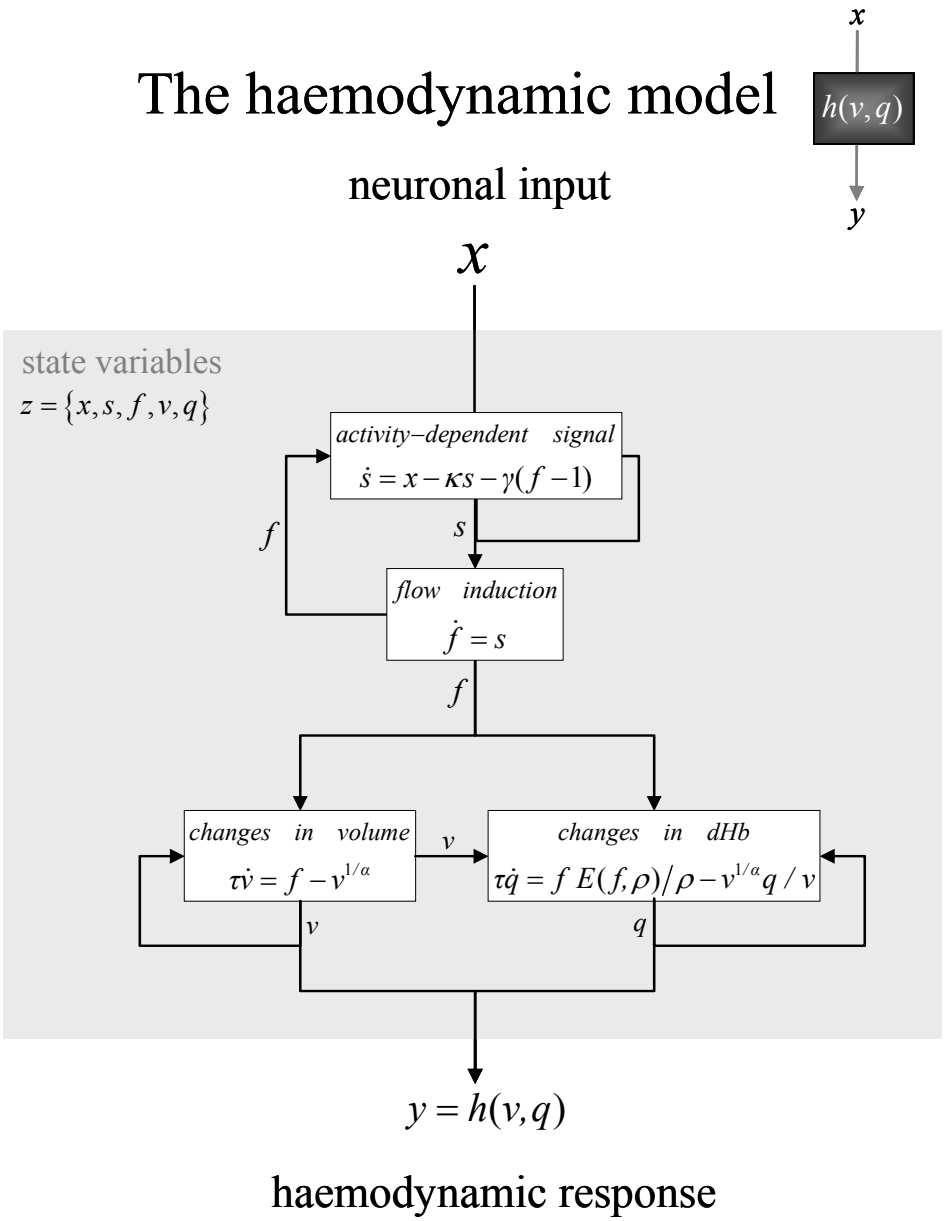


Figure 2.2: Summary of the haemodynamic model used by DCM for fMRI. Neuronal activity induces a vasodilatory and activity-dependent signal *s* that increases blood flow *f*. Blood flow causes changes in volume and deoxyhaemoglobin (*v* and *q*). These two haemodynamic states enter the output nonlinearity which results in a predicted BOLD response *y*. The model has 5 haemodynamic parameters: the rate constant of the vasodilatory signal decay (κ), the rate constant for auto-regulatory feedback by blood flow (γ), transit time (τ), Grubb's vessel stiffness exponent (α), and capillary resting net oxygen extraction (ρ). *E* is the oxygen extraction function. This figure

encodes graphically the transformation from neuronal states x_i to haemodynamic response y_i, adapted from (Friston et al., 2003).

Together, the neuronal and hemodynamic state equations yield a deterministic forward model with hidden states. For any given combination of parameters θ and inputs u, the measured BOLD response y is modelled as the predicted BOLD signal $h(u,\theta)$ plus a linear mixture of confounds $X\beta$ (e.g. signal drift) and Gaussian observation error e:

$$y = h(x,u,\theta) + \mathrm{X}\beta + \varepsilon \tag{2.6}$$

The combined neural and haemodynamic parameter set $\theta = \left\{\theta^{(n)}, \theta^{(h)}\right\}$ is estimated from the measured BOLD data, using a fully Bayesian approach with empirical priors for the haemodynamic parameters and conservative shrinkage priors for the coupling parameters. Details of the parameter estimation scheme, which rests on an expectation maximization (**EM**; see *Appendix* B and Dempster *et al* 1977) algorithm and uses a Laplace (i.e. Gaussian) approximation to the true posterior, can be found in (Friston, 2002b). Once the parameters of a DCM have been estimated from measured BOLD data, the posterior distributions of the parameter estimates can be used to test hypotheses about connection strengths. Due to the Laplace approximation, the posterior distributions are defined by their posterior mode or maximum a posteriori (MAP) estimate and their posterior covariance.

2.3 Dynamic causal modelling using neural mass models

Event-related potentials (ERPs) have been used for decades as electrophysiological correlates of perceptual and cognitive operations. However, the exact neurobiological mechanisms underlying their generation are still unclear (Baillet et al., 2001). DCM for ERPs was developed as a biologically plausible model to understand how event-related responses result from the dynamics in coupled neural ensembles. It rests on a neural mass model (NMM) which uses established connectivity rules in hierarchical

sensory systems to assemble a network of coupled cortical sources (David and Friston, 2003; David et al., 2005; Jansen and Rit, 1995).

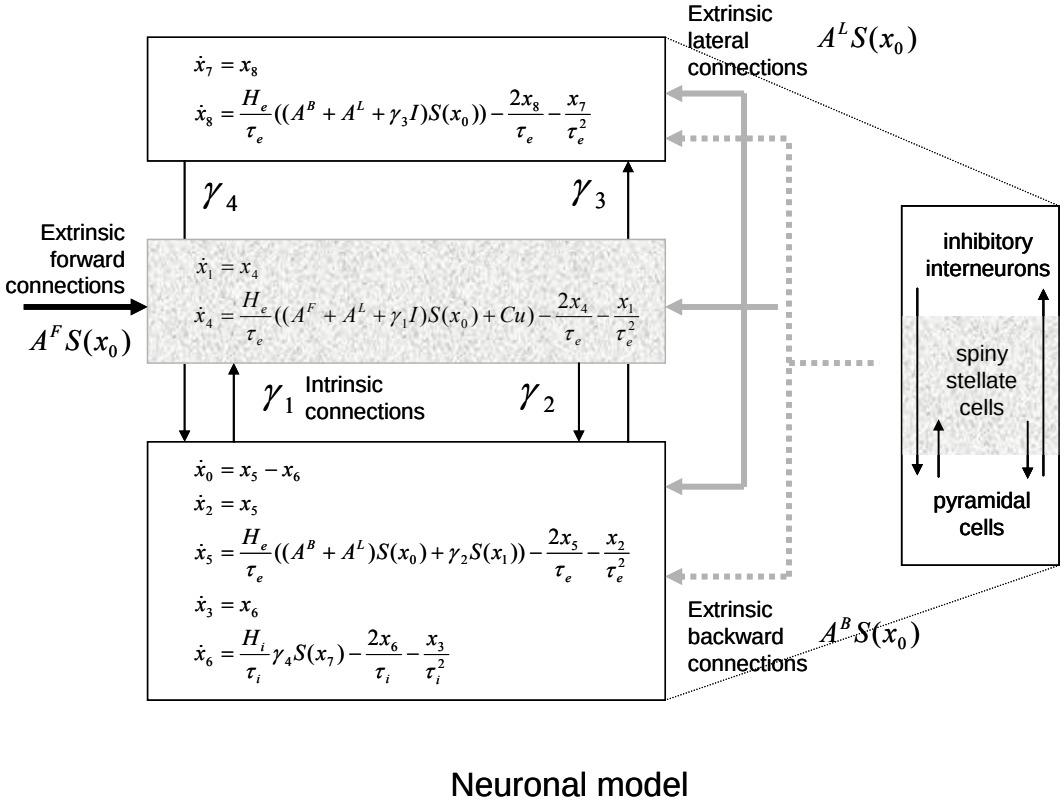


Figure 2.3: Schematic of the DCM used to model electrophysiological responses. This schematic shows the state equations describing the dynamics of sources or regions. Each source is modelled with three subpopulations (pyramidal, spiny stellate and inhibitory interneurons) as described in (Jansen and Rit, 1995) and in (David and Friston, 2003). These have been assigned to granular and agranular cortical layers which receive forward and backward connections respectively, (David et al., 2006a).

The DCM developed (David et al., 2006a), uses the connectivity rules described in (Felleman and Van Essen, 1991) to assemble a network of coupled sources. These rules are based on a partitioning of the cortical sheet into supra-, infra-granular layers and granular layer (layer 4). Bottom-up or forward connections originate in agranular layers and terminate in layer 4. Top-down or backward connections target agranular layers. Lateral connections originate in agranular layers and target all layers. These

long-range or extrinsic cortico-cortical connections are excitatory and arise from pyramidal cells.

Each region or source is modelled using a neural mass model described in (David and Friston, 2003), based on the model of (Jansen and Rit, 1995). This model emulates the activity of a cortical area using three neuronal subpopulations, assigned to granular and agranular layers. A population of excitatory pyramidal (output) cells receives inputs from inhibitory and excitatory populations of interneurons, via intrinsic connections (intrinsic connections are confined to the cortical sheet). Within this model, excitatory interneurons can be regarded as spiny stellate cells found predominantly in layer 4 and in receipt of forward connections. Excitatory pyramidal cells and inhibitory interneurons are considered to occupy agranular layers and receive backward and lateral inputs (see Figure 2.3).

To model event-related responses, the network receives inputs via input connections. These connections are exactly the same as forward connections and deliver inputs to the spiny stellate cells in layer 4. The vector C controls the influence of the input on each source. The lower, upper and leading diagonal matrices A^F, A^B, A^L encode forward, backward and lateral connections respectively. The DCM here is specified in terms of the state equations shown in Figure 2.3 and a linear output equation

$$\begin{aligned} \dot{x} &= f(x,u,\theta) \\ y &= Lx_0 + \varepsilon \end{aligned} \tag{2.7}$$

where x_0 represents the trans-membrane potential of pyramidal cells and L is a lead field matrix coupling electrical sources to the EEG channels (Kiebel et al., 2006). This should be compared to the DCM above for haemodynamics; here the equations governing the evolution of neuronal states are much more complicated and realistic, as opposed to the bilinear approximation in Eq. 2.5. Conversely, the output equation is a simple linearity, as opposed to the nonlinear observer used for fMRI. As an example, the state equation for the inhibitory subpopulation is[1]

[1] Propagation delays on the extrinsic connections have been omitted for clarity here and in Figure 2.3.

$$\begin{aligned}\dot{x}_7 &= x_8 \\ \dot{x}_8 &= \frac{H_e}{\tau_e}((A^B + A^L + \gamma_3 I)S(x_0)) - \frac{2x_8}{\tau_e} - \frac{x_7}{\tau_e^2}\end{aligned} \tag{2.8}$$

Within each subpopulation, the evolution of neuronal states rests on two operators. The first transforms the average density of pre-synaptic inputs into the average postsynaptic membrane potential. This is modelled by a linear transformation with excitatory and inhibitory kernels parameterised by $H_{e,i}$ and $\tau_{e,i}$. $H_{e,i}$ control the maximum post-synaptic potential and $\tau_{e,i}$ represent a lumped rate-constant. The second operator *S* transforms the average potential of each subpopulation into an average firing rate. This is assumed to be instantaneous and is a sigmoid function. Interactions, among the subpopulations, depend on constants $\gamma_{1,2,3,4}$, which control the strength of intrinsic connections and reflect the total number of synapses expressed by each subpopulation. In Eq. 2.8 the first line expresses the rate of change of voltage as a function of current. The second line specifies how current changes as a function of voltage, current and pre-synaptic input from extrinsic and intrinsic sources. Having specified the DCM in terms of these equations one can estimate the coupling parameters from empirical data using **EM** (see *Appendix* B). Just as with DCM for fMRI, the DCM for ERPs is usually used to investigate whether coupling strengths change as a function of experimental context.

2.4 Bayesian model selection

A generic problem encountered by any kind of modelling approach is the question of model selection: given some observed data, which of several alternative models is the optimal one? This problem is not trivial because the decision cannot be made solely by comparing the relative fit of the competing models. One also needs to take into account the relative complexity of the models as expressed, for example, by the number of free parameters in each model. Model complexity is important to consider because there is a trade-off between model fit and generalisability (i.e. how well the model explains different data sets that were all generated from the same underlying

process). As the number of free parameters is increased, model fit increases monotonically whereas beyond a certain point model generalisability decreases. The reason for this is 'overfitting': an increasingly complex model will, at some point, start to fit noise that is specific to one data set and thus become less generalisable across multiple realizations of the same underlying generative process. [Generally, in addition to the number of free parameters, the complexity of a model also depends on its functional form; see (Pitt and Myung, 2002). This is not an issue for DCM, however, because Bayesian model selection (BMS) accommodates different functional forms; see below.].

Therefore, the question "Which is the optimal model among several alternatives?" can be reformulated more precisely as "Given several alternatives, which model represents the best balance between fit and complexity?" In a Bayesian context, the latter question can be addressed by comparing the evidence, $p(y|m)$, of different models. According to Bayes theorem

$$p\left(\theta|y,m\right)=\frac{p\left(y|\theta,m\right)p\left(\theta,m\right)}{p\left(y|m\right)} \tag{2.9}$$

the model evidence can be considered as a normalization constant for the product of the likelihood of the data and the prior probability of the parameters, therefore

$$p(y\,|\,m)=\int p(y\,|\,\theta,m)\,p(\theta\,|\,m)d\theta\,. \tag{2.10}$$

Here, the number of free parameters (as well as the functional form) are considered by the integration. Unfortunately, this integral cannot usually be solved analytically, therefore an approximation to the model evidence is needed, see *Appendix* C.

In the context of DCM, one potential solution could be to make use of the Laplace approximation, i.e. to approximate the model evidence by a Gaussian that is centred on its mode. As shown by (Penny et al., 2004a), this yields the following expression for the natural logarithm (*ln*) of the model evidence ($\eta_{\theta|y}$ denotes the MAP estimate, $C_{\theta|y}$ is the posterior covariance of the parameters, C_ε is the error covariance, θ_P is the prior mean of the parameters, and C_P is the prior covariance):

$$\ln p(y \mid m) = accuracy(m) - complexity(m)$$

$$= \left[-\tfrac{1}{2} \ln \mid C_\varepsilon \mid -\tfrac{1}{2} \varepsilon_y{}^T C_\varepsilon^{-1} \varepsilon_y \right] - \left[\tfrac{1}{2} \ln \mid C_P \mid -\tfrac{1}{2} \ln \mid C_{\theta \mid y} \mid + \tfrac{1}{2} \varepsilon_\theta{}^T C_P^{-1} \varepsilon_\theta \right]. \quad (2.11)$$

$$\varepsilon_y = y - h(u, \eta_{\theta \mid y})$$
$$\varepsilon_\theta = \eta_{\theta \mid y} - \theta_P$$

This expression properly reflects the requirement, as discussed above, that the optimal model should represent the best compromise between model fit (accuracy) and model complexity. Model selection is then based on that approximation; where the best model gives the greater *Bayes factor* (BF; (Kass and Raftery, 1995)):

$$BF_{ij} = \frac{p(y \mid m_i)}{p(y \mid m_j)}. \quad (2.12)$$

Just as conventions have developed for using *p*-values in frequentist statistics, there are conventions for the use of BFs. For example, (Raftery, 1995) suggests interpretation of BFs as providing weak (BF < 3), positive (3 ≤ BF < 20), strong (20 ≤ BF < 150) or very strong (BF ≥ 150) evidence for preferring one model over another. BMS plays a central role in the application of DCM. It can be seen that the Bayes factor is the same as the difference in log-evidences. This means that the best model among competing models is the model with the greatest log-evidence. We will use log-evidence (or its free energy approximation) for BMS in the remainder of this thesis. The search for the best model, amongst several competing ones, precedes (and is often equally important to) the question which parameters of the model represent significant effects. Several studies have used BMS (Penny et al., 2004a; Stephan et al., 2007b; Stephan et al., 2009) successfully to address complex questions about the architecture of neural systems.

Comparison at the between-subject level has been used extensively in previous group studies in neuroimaging through group Bayes factor (*GBF*). For example, the *GBF* has been used frequently to decide between competing DCMs fitted to fMRI (Acs and

Greenlee, 2008; Allen et al., 2008; Grol et al., 2007; Heim et al., 2009; Kumar et al., 2007; Leff et al., 2008; Smith et al., 2006; Stephan et al., 2007c; Summerfield and Koechlin, 2008) and EEG data (Garrido et al., 2008; Garrido et al., 2007b). *GBF* is simply the product of Bayes factors over *N* subjects:

$$GBF_{i,j} = \prod_{n=1}^{N} BF_{i,j}^{(n)} \tag{2.13}$$

Here, the subscripts *i,j* refer to the models being compared, and the bracketed superscript refers to the n^{th} subject. This is equivalent to a fixed effects analysis that rests on multiplying the likelihoods over subjects to furnish the probability of the multi-subject data, conditioned on each model. This is fundamentally different from a generative model which treats subjects as random effects: here we would select a model for each subject by sampling from a multinomial distribution, and then generate data under that subject-specific model. Whenever subjects can exhibit different models or functional architectures, the random effects BMS technique presented in (Stephan et al., 2009) is a more appropriate method. In the context of basic mechanisms that are unlikely to differ across subjects, the conventional GBF is both sufficient and appropriate.

2.5 Conclusion

In this Chapter we have reviewed DCM and BMS. By creating observation models based on explicit forward models of neuronal interactions, one can model and assess interactions among distributed cortical areas and make inferences about coupling at the neuronal level. BMS has a key role in the search for the best model and in the application of DCM. The next years will probably see an increasing realism in the dynamic causal models introduced above. These endeavours are likely to encompass fMRI signals enabling the conjoint modelling, or fusion, of different modalities and the marriage of computational neuroscience with the modelling of brain responses.

CHAPTER 3

DYNAMICAL CAUSAL MODELLING FOR FMRI: A TWO-STATE MODEL

The previous Chapter presented DCM as a novel tool for modelling and analysis of connectivity in the brain. As we saw, DCM for fMRI is a technique for inferring directed connectivity among brain regions. This model distinguishes between a neuronal level, which models neuronal interactions among regions and an observation level, which models the haemodynamic responses in each region. The original DCM formulation considered only one neuronal state per region. In this Chapter, we adopt a more plausible and less constrained neuronal model, using two neuronal states (populations) per region. Critically, this gives us an explicit model of intrinsic (between-population) connectivity within a region. In addition, by using positivity constraints, the model conforms to the organisation of real cortical hierarchies, whose extrinsic connections are excitatory (glutamatergic). By incorporating two populations within each region we can model selective changes in both extrinsic and intrinsic connectivity.

Using synthetic data, we show that the two-state model is internally consistent and identifiable. We then apply the model to real data, explicitly modelling intrinsic connections. Using model comparison, we found that the two-state model is better than the single-state model. Furthermore, using the two-state model we find that it is possible to disambiguate between subtle changes in coupling; we were able to show that attentional gain, in the context of visual motion processing, is accounted for sufficiently by an increased sensitivity of excitatory populations of neurons in V5, to forward afferents from earlier visual areas.

3.1 Introduction

Dynamic Causal Modelling (DCM) for fMRI is a natural extension of the convolution models used in the standard analysis of fMRI (Friston et al., 2003). This extension involves the explicit modelling of activity within and among regions of a hypothesized network, at the neuronal level. The general idea behind DCM is to construct a reasonably realistic neuronal model of interacting cortical regions with neurophysiologically inspired parameters. These parameters are estimated such that the predicted blood oxygenation level dependent (BOLD) series, which results from converting the neural dynamics into haemodynamics, correspond as closely as possible to the observed BOLD series.

Standard DCMs for fMRI are based upon a bilinear approximation to neuronal dynamics with one state per region. The neuronal dynamics are described by the differential equations describing the dynamics of a single state that summarises the neuronal or synaptic activity of each area; this activity then induces a haemodynamic response as described by an extended Balloon model (Buxton et al., 1998). Examples of DCM for fMRI can be found in (Bitan et al., 2005; den Ouden et al., 2008; Eickhoff et al., 2008; Ethofer et al., 2006; Fairhall and Ishai, 2007; Griffiths et al., 2007; Kumar et al., 2007; Leff et al., 2008; Mechelli et al., 2005; Mechelli et al., 2004; Mechelli et al., 2003; Noppeney et al., 2006; Posner et al., 2006; Stephan et al., 2006; Stephan et al., 2007b; Stephan et al., 2005; Summerfield et al., 2006). For a review on the conceptual basis of DCM and its implementation for functional magnetic resonance imaging data and event-related potentials see (Stephan et al., 2007a).

Dynamical Causal Modelling differs from established methods for estimating effective connectivity from neurophysiological time series, which include structural equation modelling and models based on multivariate autoregressive processes (Harrison et al., 2003; McIntosh and Gonzalez-Lima, 1994; Penny et al., 2004b; Roebroeck et al., 2005). In these models, there is no designed perturbation and the inputs are treated as unknown and stochastic. DCM assumes the input to be known, which seems appropriate for designed experiments. Further, DCM is based on a

parameterised set of differential equations which can be extended to better describe the system.

Here, we extend the original model to cover two states per region. These states model the activity of inhibitory and excitatory populations. This has a number of key advantages. First, we can relax any *shrinkage priors* used to enforce stability in single-state DCMs, because the interaction of excitatory-inhibitory pairs confers dynamical stability[2] on the system. Second, we can model both extrinsic and intrinsic connections. Third, we can enforce positivity constraints on the extrinsic connections (*i.e.*, inter-regional influences of excitatory populations). Finally, this re-parameterisation enables one to model context-dependent changes in coupling as a proportional increase or decrease in connection strength (*c.f.*, the additive effects used previously, (Friston et al., 2003)).

Shrinkage priors are simply priors or constraints on the parameters that shrink their conditional estimates towards zero (i.e., their prior expectation is zero and the prior variance determines the degree of shrinkage, in relation to observation noise). They were employed in early formulations of DCM to ensure coupling strengths did not attain very high weights, which generate exponentially diverging neuronal activity. However, this motivation for shrinkage priors is rather *ad hoc* and, as we will discuss later, confounds model specification and comparison.

This Chapter is structured as follows. In the first section, we present the two-state DCM, with two states per region. In the subsequent section, we provide a stability analysis of the two-state DCM. In the third section, we describe model inversion; i.e., prior distributions, Bayesian estimation, conditional inference and model comparison. In section four, we compare the single and two-state DCM using synthetic and real data to establish its face validity. Finally, an empirical section then demonstrates the use of the two-state DCM by looking at attentional modulation of connections during visual motion processing. From these analyses, we conclude that the two-state DCM is a better model for fMRI data than the single-state DCM.

[2] Excitatory-Inhibitory models are not generically stable in a dynamical sense. However, note that this is probably not an issue during inversion, because we never found that the inversion step identifies parameters that lead to instable behaviour.

3.2 Theory

3.2.1 Dynamic Causal Modelling for fMRI – Single-state models

In this section we review briefly dynamic causal models of fMRI data (Chapter 2; (Friston et al., 2003)). In the next section, we extend this model to accommodate two neuronal sources per region. In dynamic causal models, interactions among regions are modelled at the neuronal level. In single-state models each region has one state variable. This state is a simple summary of neuronal (i.e., synaptic) activity $x(t)$, in a region. (Friston et al., 2003) used a bilinear form to describe their dynamics:

$$\begin{aligned} \dot{z} &= F(x,u,\theta) \approx \Im x + Cu \\ \Im &= A + \sum_j u_j B^{(j)} \end{aligned}$$

$$\begin{aligned} A &= \frac{\partial F}{\partial x} = \left.\frac{\partial \dot{x}}{\partial z}\right|_{u=0} \\ B^{(j)} &= \frac{\partial^2 F}{\partial x \partial u_j} = \frac{\partial}{\partial u_j}\frac{\partial \dot{x}}{\partial z} \\ C &= \left.\frac{\partial F}{\partial u}\right|_{x=0} \end{aligned} \quad . \tag{3.1}$$

This model is used to generate neuronal activity; later we will add haemodynamics and noise to furnish a probabilistic model of fMRI measurements. In this model, the state vector $x(t)$ contains one scalar per region. The changes in neuronal (i.e., synaptic) activity are described by the sum of three effects. First, the matrix A encodes directed connectivity between pairs of regions. The elements of this connectivity matrix are not a function of the input, and can be considered as an endogenous or condition-invariant. Second, the elements of $B^{(j)}$ represent the changes of connectivity induced by the inputs, u_j. These condition-specific modulations or bilinear terms $B^{(j)}$ are usually the interesting parameters. The endogenous and condition-specific matrices are mixed to form the total connectivity

or Jacobian matrix $\Im$. Third, there is a direct exogenous influence of each input u_j on each area, encoded by the matrix C. The parameters of this system, at the neuronal level, are given by $\theta^n \supseteq A, B^1, \ldots, B^{N_u}, C$. At this level, one can specify which connections one wants to include in the model. Connections (*i.e.*, elements of the matrices) are removed by setting their prior mean and variance to zero. We will illustrate this later.

The bilinear form in Eq. 3.1 can be regarded as an approximation to any function, $F(x,u,\theta)$, because it is simply a Taylor expansion around $x = 0$ and $u = 0$; retaining only terms that are first-order in the states or input. In this sense, the bilinear model can be regarded as a generic approximation, to any [unknown] function describing neuronal dynamics, in the vicinity of its fixed-point; i.e., when the neuronal states are at equilibrium or zero.

At the observation level; for each region, the neuronal state forms an input to a haemodynamic model that generates the BOLD signal, see Chapter 2 for details.

3.2.2 Dynamic Causal Modelling for fMRI – Two-state models

We now extend the standard DCM above to incorporate two state variables per region. These model the activity of an inhibitory and excitatory population respectively. Schematics of the single and two-state models are shown in Figure 3.1.

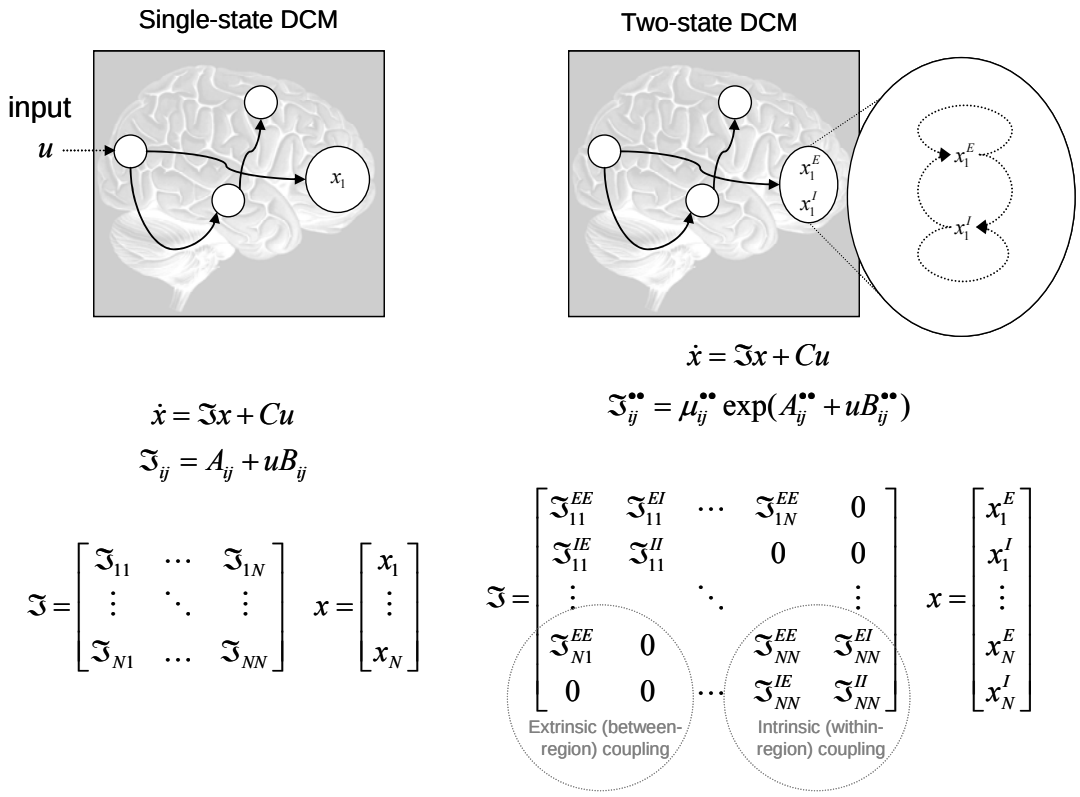


Figure 3.1: Schematic of the **Single-state DCM** (left) and the present **Two-state DCM** (right). The Two-state model has an **inhibitory** and an **excitatory** subpopulation. The positivity constraints are explicitly represented in the two-state connectivity matrix by exponentiation of underlying scale parameters (bottom right).

The Jacobian matrix, $\Im$ represents the effective connectivity within and between regions. *Intrinsic* or within-region coupling is encoded by the leading diagonal blocks (see Figure 3.1), and *extrinsic* or between-region coupling is encoded by the off-diagonal blocks. Each within-region block has four entries, $\Im_{ii}^{\bullet\bullet} = \{\Im_{ii}^{EE}, \Im_{ii}^{II}, \Im_{ii}^{EI} \Im_{ii}^{IE}\}$. These correspond to all possible intrinsic connections between the excitatory and inhibitory states, $\{x_i^E, x_i^I\}$ of the *i*-th region. These comprise self-connections, $E \to E$, $I \to I$ and inter-state connections $E \to I$, $I \to E$. We enforce the connections, $E \to E$, $I \to E$, $I \to I$ to be negative (*i.e.*, $\Im_{ii}^{EE}, \Im_{ii}^{IE}, \Im_{ii}^{II} \le 0$), which means they mediate a dampening effect on population responses. This negativity is imposed by using log-normal priors; we use the negative exponential of an underlying coupling parameter with a normal prior (see below). Although the excitatory self-

connections are negative, we do not mean to suggest that there are direct inhibitory connections among excitatory units; rather the multitude of mechanisms that self-organise neuronal activity (*e.g.*, adaptation, gain-control, refractoriness, polysynaptic input from recurrent axonal collaterals, *etc*.) will conspire to make the effective self-connection negative[3]. The extrinsic connections among areas are assumed to be positive (*i.e.*, $\Im_{ij}^{EE} \geq 0$) and are mediated exclusively by coupling among excitatory populations (c.f., glutamatergic projections in the real brain). In accord with known anatomy, we disallow long-range coupling among inhibitory populations.

The two-state DCM has some significant advantages over the standard DCM. First, intrinsic coupling consists of excitatory and inhibitory influences, which is biologically more plausible. Also, the interactions between inhibitory and excitatory subpopulations confer more stability on the overall system. This means we can relax the shrinkage priors used to enforce stability in single-state DCMs. Furthermore, we can now enforce positivity constraints on the extrinsic connections (*i.e.*, inter-regional influences among excitatory populations) using log-normal priors and scale parameters as above for the intrinsic connections. This means changes in connectivity are now expressed as a proportional increase or decrease in connection strength. In what follows, we address each of these issues, starting with the structural stability of two-state systems and the implications for priors on their parameters.

3.3 Stability and priors

In this section, we will describe a stability analysis of the two-state system, which informs the specification of the prior distributions of the parameters. Network models like ours can display a variety of different behaviours (e.g., (Wilson and Cowan, 1972, 1973). This is what makes them so useful, but there are parameterizations which make the system unstable. By this, we mean that the system response increases exponentially. In real brains, such behaviour is not possible and this domain of parameter space is highly unlikely to be populated by neuronal systems. The prior

[3] In the present described DCM extension, we used the original adopted self-dampening E→E connection (Friston et al. (2003). Though, self-regularization mechanism could possibly also be obtained by exclusively inhibitory I→E connection parameterizations.

distributions should reflect this by assigning a prior probability of zero to unstable domains. However, this is not possible, because we have to use normal priors to keep the model inversion analytically tractable. Instead, we specify priors that are centred on stable regions of parameter space.

In the original single-state DCM, we had a single state per region and a self-decay for each state (see Figure 3.2). This kind of system allows for only an exponential decay of activity in each region, following a perturbation of the state by exogenous input or incoming connections. For this model, (Friston et al., 2003) chose shrinkage priors, which were used to initialise the inversion scheme in a stable regime of parameter space, in which neuronal activity decayed rapidly. The conditional parameter estimates were then guaranteed to remain in a stable regime through suitable checks during iterative optimisation of the parameters.

Figure 3.2: Schematic of **Single-state DCM** (one region).

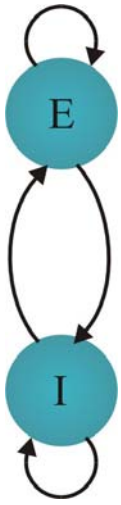


Figure 3.3: Schematic of **Two-state DCM** (one region).

Two-state models (Figure 3.3) can exhibit much richer dynamics compared to single-state models. One can determine analytically the different kinds of periodic and harmonic oscillatory network modes these systems exhibit (see *Appendix* D). This is important because it enables us to establish stability for any prior mean of the parameters. This entails performing a linear stability analysis by examining the eigenvalues of the Jacobian, $\Im$ under the prior expectation of the parameters. The system is asymptotically stable if these eigenvalues (c.f., Lyapunov spectrum) have only negative real parts (Dayan, 2001). This is the procedure we adopt below.

It should be noted that, in generic coupled nonlinear systems, instability of a linearly stable fixed point does not always lead to exponential growth, but may lead to the appearance of a stable nonlinear regime. In the case of a Hopf bifurcation (as in (Wilson and Cowan, 1973), a limit cycle appears near the unstable fixed point, which can model alpha rhythms and other oscillatory phenomena. Indeed, a system close to a linear instability exhibits longer and more complex nonlinear transients on perturbation (e.g., (Friston, 1997)). This is a further reason to avoid using shrinkage priors (that preclude systems close to instability). However, because the bilinear model is linear in its states, its unstable fixed points are necessarily associated with exponential growth.

3.3.1 Priors

We now describe how we specify the priors and enforce positivity or negativity constraints on the connections. We seek priors that are specified easily and are not a function of connectivity structure; because this can confound model comparison (Penny et al., 2004a). The strategy we use is to determine a stable parameterization for a single area, use this for all areas and allow only moderate extrinsic connections. In this way, the system remains stable for all plausible network structures.

Priors have a dramatic impact on the landscape of the objective function that is optimised: good choices of prior distributions will help to reach the appropriate posterior distributions by means of identifying the global minimum of the objective

function. Under Gaussian assumptions, the prior distribution $p(\theta)$ is defined by its mean and covariance Σ. Since Gaussian priors have infinite support, we will always have finite density in unstable areas. However, we evaluate the eigenvalues of the Jacobian to start the estimation in a stable domain and we observe experimentally that with priors that have the most of their mass in stable areas, the data will support posterior means in stable areas. In our expectation-maximization (EM) inversion scheme, the prior expectation is also the starting estimate. If we chose a stable prior, we are guaranteed to start in a stable domain of parameter space. After this initialization, the algorithm could of course update to an unstable parameterization, because we are dealing with a dynamic generative model. However, these updates will be rejected because they cannot increase the objective function: in the rare cases when an update to an unstable regime actually occurs (and the objective function decreases), the algorithm returns to the previous estimate and halves its step-size, using a Levenberg-Marquardt scheme (Press, 1999). This is repeated iteratively, until the objective function increases; at which point the update is accepted and the optimization proceeds. Therefore, it is sufficient to select priors whose mean lies in a stable domain of parameter space.

Stability is conferred by enforcing connectivity parameters to be strictly positive or negative. In particular, the intrinsic, $E \to E$, $I \to I$, $I \to E$ connections are negative while the $E \to I$ and all extrinsic $E \to E$ connections are positive. We use

$$\begin{bmatrix} -1 & -0.5 \\ 0.5 & -1 \end{bmatrix}$$

as the prior mode (most likely *a priori*) for a single region's Jacobian, where its states, $x = [x_i^E x_i^I]^T$ summarise the activity of its constituent excitatory and inhibitory populations. This Jacobian has eigenvalues of $-1 \pm 0.5i$ and is guaranteed to be stable. We then replicate these priors over regions, assuming weak positive excitatory extrinsic $E \to E$ connections, with a prior of 0.5 Hz. For example, a three region model, with hierarchical reciprocal extrinsic connections and states, $x = [x_1^E, x_1^I, x_2^E, x_2^I, x_3^E, x_3^I]^T$ would have Jacobian with a prior mode of

$$\mu = \begin{bmatrix} -1 & -0.5 & 0.5 & 0 & 0 & 0 \\ 0.5 & -1 & 0 & 0 & 0 & 0 \\ 0.5 & 0 & -1 & -0.5 & 0.5 & 0 \\ 0 & 0 & 0.5 & -1 & 0 & 0 \\ 0 & 0 & 0.5 & 0 & -1 & -0.5 \\ 0 & 0 & 0 & 0 & 0.5 & -1 \end{bmatrix}.$$

The eigenvalue spectrum of this Jacobian is show in Figure 3.4 (left panel), along with the associated impulse response functions for an input to the first subpopulation (right panels); evaluated with $x(t) = \exp(\mu t)x(0)$. It can be seen for this architecture we expect neuronal dynamics to play out over a time-scale of about one second. Note that these dynamics are not enforced; they are simply the most likely *a priori*.

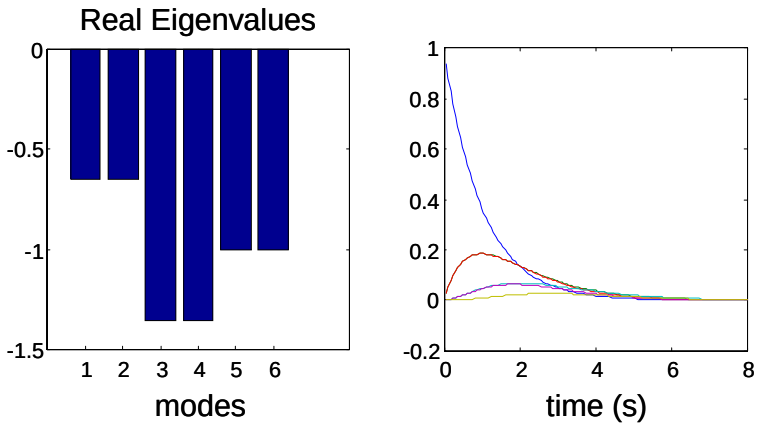


Figure 3.4: Stability analyses for a two-state DCM for three interconnected regions (see Jacobian above). Left panel: Real negative (stable) eigenmodes. Right panel: Associated impulse response functions evaluated with $x(t) = \exp(\mu t)x(0)$.

3.3.2 Positivity constrains and scale-parameters

To ensure positivity or negativity, we scale these prior modes, μ with scale-parameters, which have log-normal priors. This is implemented using underlying coupling parameters with Gaussian or normal priors; for example, the extrinsic connections are parameterized as $\Im_{ij}^{\bullet\bullet} = \mu_{ij}^{\bullet\bullet} \exp(A_{ij}^{\bullet\bullet} + uB_{ij}^{\bullet\bullet})$, where

$\Im_{ii}^{\bullet\bullet} = \{\Im_{ii}^{EE}, \Im_{ii}^{II}, \Im_{ii}^{EI} \Im_{ii}^{IE}\}$, $p(A_{ij}^{\bullet\bullet}) = N(0, v)$ and we have assumed one input. A mildly informative log-normal prior obtains when the prior variance $v \approx 1/16$. This allows for a scaling around the prior mode, $\mu_{ij}^{\bullet\bullet}$ of up to a factor of two, where the sign of the mode determines whether the connection is positive or negative. In what follows, we use a prior variance for the endogenous and condition-specific coupling parameters, A_{ij}^{jk} and B_{ij}^{jk} of $v = 1/16$.

Re-parameterising the system in terms of scale-parameters entails a new state equation (see Figure 3.1), which replaces the Bilinear model in Eq. 3.1

$$\dot{x} = \Im x + Cu$$

$$\Im_{ij}^{\bullet\bullet} = \mu_{ij}^{\bullet\bullet} \exp(A_{ij}^{\bullet\bullet} + \sum_k u_k B_{ij}^{\bullet\bullet(k)}) = \mu_{ij}^{\bullet\bullet} \exp(A_{ij}^{\bullet\bullet}) \prod_k \exp(u_k B_{ij}^{\bullet\bullet(k)})$$

$$\Im = \begin{bmatrix} \Im_{11}^{EE} & \Im_{11}^{EI} & \cdots & \Im_{1N}^{EE} & 0 \\ \Im_{11}^{IE} & \Im_{11}^{II} & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ \Im_{N1}^{EE} & 0 & & \Im_{NN}^{EE} & \Im_{NN}^{EE} \\ 0 & 0 & \cdots & \Im_{NN}^{IE} & \Im_{NN}^{II} \end{bmatrix} \quad x = \begin{bmatrix} x_1^E \\ x_1^I \\ \vdots \\ x_N^E \\ x_N^I \end{bmatrix} \quad . \tag{3.2}$$

In this form, it can be seen that condition-specific effects u_k act to scale the connections by $\exp(u_k B_{ij}^{\bullet\bullet(k)}) = \exp(B_{ij}^{\bullet\bullet(k)})^{u_k}$. When $B_{ij}^{\bullet\bullet(k)} = 0$, this scaling is $\exp(u_k B_{ij}^{\bullet\bullet(k)}) = 1$ and there is no effect of input on the connection strength. The haemodynamic priors are those used in (Friston, 2002b).

Having specified the form of the DCM in terms of its likelihood and priors, we can now estimate its unknown parameters, which represent a summary of the coupling among brain regions and how they change under different experimental conditions.

3.4 Bayesian estimation, inference and model comparison

For a given DCM, say model m, parameter estimation corresponds to approximating the moments of the posterior distribution given by Bayes rule

$$p(\theta \mid y,m) = \frac{p(y \mid \theta,m)\,p(\theta \mid m)}{p(y \mid m)}. \tag{3.3}$$

The estimation procedure employed in DCM is described in (Friston et al., 2003; Kiebel et al., 2006). The posterior moments (conditional mean η and covariance Ω) are updated iteratively using an expectation maximization (EM; see *Appendix* B and Dempster *et al* 1977), which uses a fixed-form Laplace (*i.e.*, Gaussian) approximation to the conditional density $q(\theta) = N(\eta,\Omega)$.

Often, one wants to compare different models for a given data set. We use Bayesian model comparison, using the model evidence (Penny et al., 2004a), which is

$$p(y \mid m) = \int p(y \mid \theta,m)\,p(\theta \mid m)d\theta. \tag{3.4}$$

Note that the model evidence is the normalization constant in Eq. 3.3. The evidence can be decomposed into two components: an accuracy term, which quantifies the data fit, and a complexity term, which penalizes models with redundant parameters. In the following, we approximate the model evidence for model *m*, under the Laplace approximation, with

$$\ln p(y \mid m) \approx \ln p(y \mid \lambda,m), \tag{3.5}$$

where λ are the unknown covariance component parameters, i.e., the hyperparameters (Friston et al., 2007). This is the maximum value of the objective function attained by **EM**. The most likely model is the one with the largest log-evidence. This enables

BMS. Model comparison rests on the likelihood ratio of the evidence for two models. This ratio is the Bayes factor B_{ij}. For models *i* and *j*

$$\ln B_{ij} = \ln p(y|m = i) - \ln p(y|m = j) . \quad (3.6)$$

Conventionally, strong evidence in favour of one model requires the difference in log-evidence to be about three or more (Penny et al., 2004a). Under the assumption that all models are equally likely *a priori*, the marginal densities $p(y \mid m)$ can be converted into the probability of the model given the data $p(m \mid y)$ (by normalising so that they sum to one over models). We will use this probability to quantify model comparisons below (see Tables).

3.5 Simulations –models comparisons

Here, we establish the face validity of the DCM described in the previous section. This was addressed by integrating DCMs with known parameters, adding observation noise to simulate responses and inverting the models. Crucially, we used different models during both generation and inversion and evaluated all combinations to ensure that model selection identified the correct model.

The DCMs used the posterior or conditional means from three different models estimated using real data (see next section). We added random noise such that the final data had a signal-to-noise ratio of three, which corresponds to typical DCM data[4]. We created three different synthetic data sets corresponding to a forward, backward and intrinsic model of attentional modulation of connections in the visual processing stream. We used a hierarchal three-region model where stimulus-bound visual input entered at the first or lowest region. In the forward model, attention increased coupling in the extrinsic forward connection to the middle region; in the backward model it changed backward influences on the middle region and in the intrinsic model attention changed the intrinsic $I \rightarrow E$ connection. In all models,

[4] Note that a DCM time-series of a single region is the first eigenvariate of a cluster of voxels and is relatively de-noised.

attention increased the sensitivity of the same excitatory population to different sorts of afferents.

We then used the three models to fit each of these three synthetic data sets, giving nine model inversions. Table 3.1 presents the log-evidences for each inversion. The highest evidence, was obtained for models that were used to generate the synthetic data: these correspond to the diagonal entries. These results show that model comparison can identify reliably the correct model, among competing and subtly different two-state models.

Models	***Synthetic Data***		
	Backward	**Forward**	**Intrinsic**
Backward	**524** **99.9%**	494 0.0%	478 0.7%
Forward	382 0.0%	**538** **99.9%**	-439 0.0%
Intrinsic	497 0.0%	504 0.0%	**482** **99.3%**

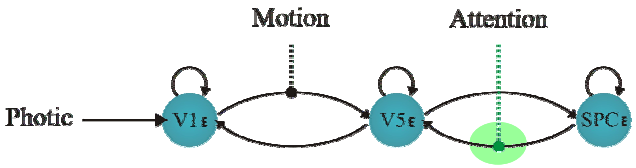
Table 3.1: Log-evidences for three different models using synthetic data generated by the Backward, Forward and Intrinsic models (see text). The diagonal values show higher log evidences, which indicate that the two-state DCM has internal consistency. The percentages correspond to the conditional probability of each model, assuming uniform priors over the three models examined under each data set.

3.6 Empirical analysis - models comparisons

In this section, we ask whether the two-state extension described in this Chapter is warranted, in terms of providing a better explanation of real data. This was addressed by inverting the single- and two-state models using the same empirical data. These data have been used previously to validate DCM and are available from http://www.fil.ion.ucl.ac.uk/spm. We analysed data from a study of attentional modulation during *visual motion processing* (Büchel and Friston, 1997). The experimental manipulations were encoded as three exogenous inputs: A '**photic stimulation**' input indicated when dots were presented on a screen, a '**motion**' variable indicated that the dots were moving and the '**attention**' variable indicated that the subject was attending to possible velocity changes. The activity was modelled in three regions **V1**, **V5** and superior parietal cortex (**SPC**).

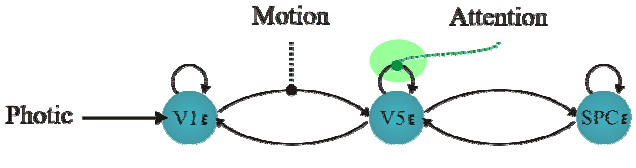
We compared the single- and two-state DCM over the following three model variants. **Model 1** assumed that attention modulates the backward extrinsic connection from **SPC** to **V5**. **Model 2** assumed that attention modulates the intrinsic connection in **V5** and **Model 3** assumed attention modulates the forward connection from **V1** to **V5**. All models assumed that the effect of motion was to modulate the connection from **V1** to **V5.** In Figure 3.5 we show each of these three variants for the single- and two-state DCM.

Model 1



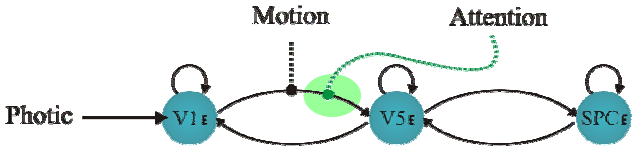
$$\begin{bmatrix} \Im^{EE}_{V1V1} & \Im^{EI}_{V1V1} & 0 \\ \Im^{EE}_{V5V1} & \Im^{EE}_{V5V5} & \Im^{EE}_{V5SP} \\ 0 & \Im^{EE}_{SPV5} & \Im^{EE}_{SPSP} \end{bmatrix}$$

Model 2



$$\begin{bmatrix} \Im^{EE}_{V1V1} & \Im^{EI}_{V1V1} & 0 \\ \Im^{EE}_{V5V1} & \Im^{EE}_{V5V5} & \Im^{EE}_{V5SP} \\ 0 & \Im^{EE}_{SPV5} & \Im^{EE}_{SPSP} \end{bmatrix}$$

Model 3



$$\begin{bmatrix} \Im^{EE}_{V1V1} & \Im^{EI}_{V1V1} & 0 \\ \Im^{EE}_{V5V1} & \Im^{EE}_{V5V5} & \Im^{EE}_{V5SP} \\ 0 & \Im^{EE}_{SPV5} & \Im^{EE}_{SPSP} \end{bmatrix}$$

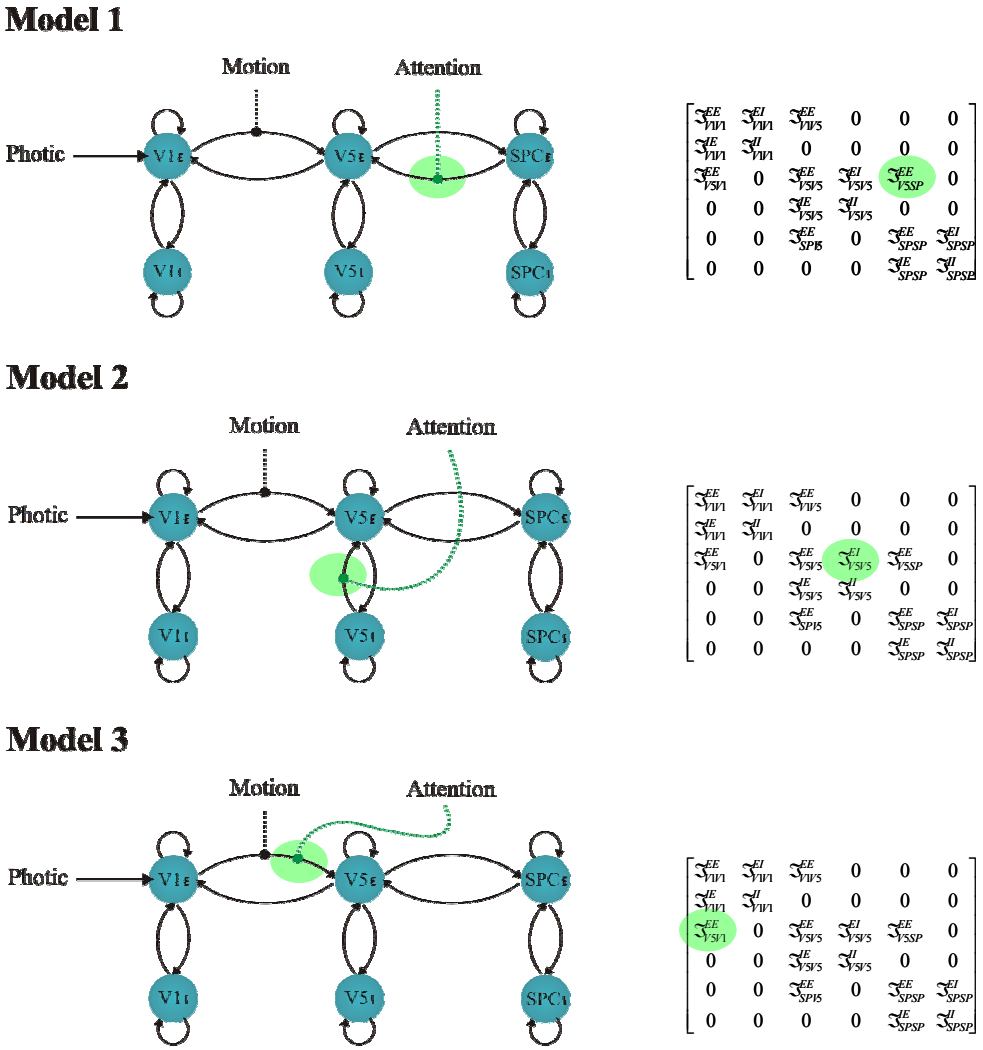


Figure 3.5: In all models photic stimulation enters **V1** and the motion variable modulates the connection from **V1** to **V5**. Models 1, 2 and 3 all assume reciprocally and hierarchically organised connections. They differ in how attention modulates the influences on **V5**; model 1 assumes modulation of the backward extrinsic connection, model 2 assumes modulation of intrinsic connections in **V5** and model 3 assumes modulation of the forward connection. **A:** single-state DCMs. **B:** two-state DCMs.

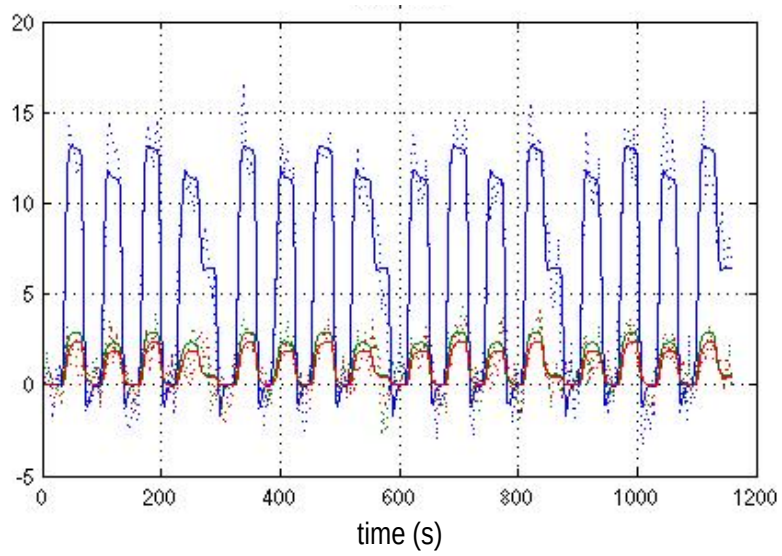


Figure 3.6: Plot of the DCM fit to visual attention fMRI data, using the two-state model 3. Solid: Prediction, Dotted: Data. Blue: **V1** response, Green: **V5** response, Red: **SPC** response.

We inverted all models using the variational **EM** scheme (see Appendix B) and compared all six DCMs using Bayesian model comparison. As a representative example of the accuracy of the DCM predictions, we show the predicted and observed BOLD series for model 3 (two-state) in Figure 3.6. The results of the Bayesian model comparison are shown in Figure 3.7, in terms of the log-evidences (in relation to a baseline model with no attentional modulation). These results show two things. First, both models find strong evidence in favour of model 3, i.e., attention modulates the forward connection from **V1** to **V5**. Second, there is strong evidence that the two-state models 2 and 3 are better than any single-state model. The respective log-evidences for this Bayesian model comparison are shown in Table 3.2. Again, the table shows that the forward model is the best model, among either the single- or two-state DCMs. Moreover, there is very strong evidence in favour of the two-state model over the single-state model, because the differences in log-evidences are all greater than five; recall that a difference in log-evidence of three or more corresponds to a Bayes factor of about 20 or more, and represents strong evidence. For reference; the log-evidence for the baseline model with no attentional modulation was -1649.9.

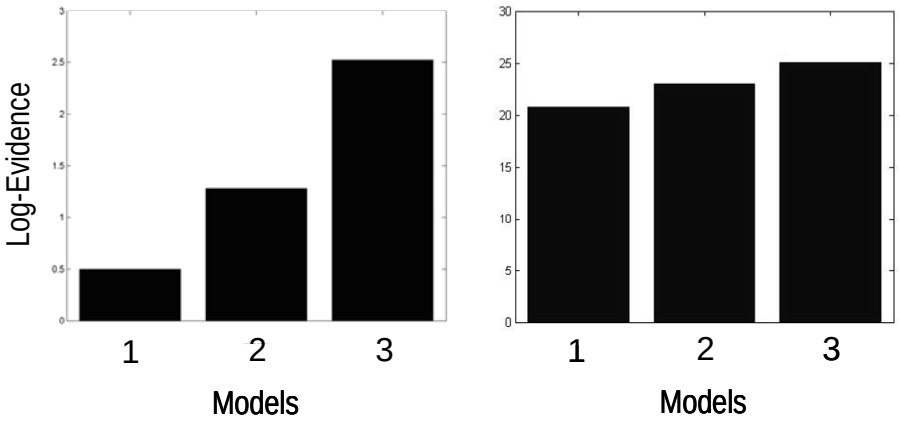


Figure 3.7: Results of the Bayesian model comparisons among DCMs for single-state (left) and two-state (right) formulations. The graphs show the log-evidences for each model (relative to a no attentional modulation model): **Model 3** (modulation of the forward connections by attention) is superior to the other two models. The two-state model log-evidences are better than any single-state model (note the difference in scale).

	Backward	**Forward**	**Intrinsic**
Single-State DCM	-1649.38 **0.00%**	-1647.36 **0.00%**	-1648.60 **0.00%**
Two-State DCM	-1629.20 **1.08%**	-1624.80 **88.12%**	-1626.90 **10.79%**
Difference in Log-Evidence	20.18	22.56	21.70

Table 3.2: This table shows the log-evidences for the two models, single and two-state DCMs, plotted in the figure 3.7. Forward modulation is the best for both models. We can also see that that there is very strong evidence in favour of the two-state model over the single-state model. The percentages in bold correspond to the

conditional probability of each model, given the data and assuming uniform priors over the six models examined.

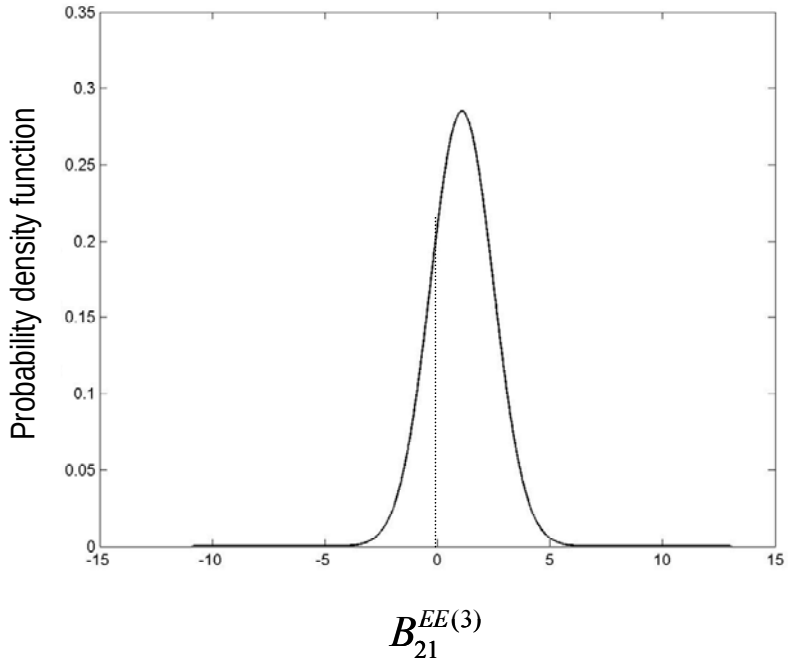


Figure 3.8: Posterior probability density functions for the Gaussian parameter, $B_{21}^{EE(3)}$ associated with attentional modulation of the forward connection in the best model. There is an 88% confidence that this gain is greater than one (area under the Gaussian to the right of the dashed line). The dashed line indicates $B_{21}^{EE(3)} = 0 \Rightarrow \exp\left(B_{21}^{EE(3)}\right) = 1$.

These results represent an inference on model space. To illustrate inference on parameter space, Figure 3.8 shows the conditional density of the parameters representing attentional gain of the forward connection in the best model. We show this conditional density on the Gaussian parameter, $B_{21}^{EE(3)}$ (with an implicit gain or scale-parameter $\exp(B_{21}^{EE(3)})$) associated with attention (*i.e.*, when $u_3 = 1$). It can be seen that we can be 88% confident that this gain is greater than one.

3.7 Discussion

In this Chapter, we have described a new DCM for fMRI, which has two states per region instead of one. With the two-state DCM, it is possible to relax shrinkage priors used to guarantee stability in single-state DCMs. Moreover, we can model both extrinsic and intrinsic connections, as well as enforce positivity constraints on the extrinsic connections.

Using synthetic data, we have shown that the two-state model has internal consistency. We have also applied the model to real data, explicitly modelling intrinsic connections. Using model comparison, we found that the two-state model is better than the single-state model and that it is possible to disambiguate between subtle changes in coupling; in the example presented here, we were able to show that attentional gain, in the context of visual motion processing, is accounted for sufficiently by an increased sensitivity of excitatory populations of neurons in V5 to forward afferents from earlier visual areas.

These results suggest that the parameterization of the standard single-state DCM is probably too constrained. With a two-state model, the data can be explained by richer dynamics at the neuronal level. This might be seen as surprising, because it generally is thought that the haemodynamic response function removes a lot of information and a reconstruction of neuronal processes is not possible. However, our results challenge this assumption, *i.e.,* DCMs with richer dynamics (and more parameters) are clearly supported by the data.

In the following, we discuss some potential extensions to current DCMs that may allow useful questions to be addressed to fMRI data: currently, we model excitatory (glutamatergic) and inhibitory (GABA-ergic) connections. As a natural extension we can include further states per region, accounting for other neurotransmitter effects. Important examples here would be adaptation phenomena and activity-dependent effects of the sort mediated by NMDA receptors. This is interesting because NMDA receptors are thought to be targeted preferentially by backward connections. This

could be tested empirically using a suitable multi-state DCM based on an explicit neural-mass model.

Another important point is that the haemodynamics in the current DCM are a function of the excitatory states only. The contributions to the BOLD signal from the inhibitory states are expressed indirectly, through dynamic interactions between the two states, at the neuronal level. One possible extension would be to model directly separate contributions of these two states, at the haemodynamic level. Hypotheses about the influence of excitatory and inhibitory populations on the BOLD signal could then be tested using model comparison.

Another extension is to generalize the interactions between the two subpopulations, i.e., to use nonlinear functions of the states in the DCM. Currently, this is purely linear in the states, but one could use sigmoidal functions. This would take our model into the class described by (Wilson and Cowan, 1973). In this fashion, one can construct more biologically constrained response functions and bring DCMs for fMRI closer to those being developed for EEG and MEG. Again, the question of whether fMRI data can inform such neural-mass models can be answered simply by model comparison. As noted above the bilinear approximation used in the original formulation of DCM for fMRI represents a global linearization over the whole of state-space; the current extension uses the same bilinear approximation in the states (although it is nonlinear in the parameters). A sigmoid nonlinearity would give a state equation that is nonlinear in the states. In this instance, we can adopt a local linearization, when integrating the system to generate predictions. In fact, our inversion scheme already uses a local linearization, because the haemodynamic part of DCM for fMRI is nonlinear in the haemodynamic states (Friston, 2002b). However, this approach does not account for noise on the states (i.e., random fluctuations in neuronal activity). There has already been much progress in the solution of stochastic differential equations entailed by stochastic DCMs, particularly in the context of neural mass models (see (Sotero et al., 2007; Valdes et al., 1999)).

Finally, in the next development of DCM for fMRI we could evaluate DCMs based on density-dynamics (see next chapter). Current DCMs consider only the mean neuronal state for each population.

3.8 Conclusion

Our results indicate that one can estimate intrinsic connection strengths within network models, using fMRI. Using real data, we find that a two-state DCM is better than the conventional single-state DCM. This demonstrates the potential of adopting generative models for fMRI time-series that are informed by anatomical and physiological principles.

CHAPTER 4

POPULATION DYNAMICS: VARIANCE AND THE SIGMOID ACTIVATION FUNCTION

The previous Chapter presented a novel DCM for fMRI, which considered two states per region, an excitatory and an inhibitory state, motivated by anatomical and physiological principles. This Chapter addresses the role of dispersion (variance) of neuronal states on the cortical responses to sensory inputs. It provides a link between the sigmoid activation function and the variance of neuronal membrane depolarization, through a cumulative density function within a population. This Chapter lays the ground-work for the following Chapters, whereby the variance itself becomes a time-dependent variable and hence dynamically coupled to the mean. This provides a crucial link between neural mass models and more general neural density models.

4.1 Introduction

The aim of this Chapter is to show how the sigmoid activation function in neural-mass models can be understood in terms of the dispersion of underlying neuronal states. Furthermore, we show how this relationship can be used to estimate the probability density of neuronal states using non-invasive electrophysiological measures such as the electroencephalogram (EEG).

There is growing interest in the use of mean-field and neural-mass models as observation models for empirical neurophysiological time-series (Breakspear et al., 2006; David and Friston, 2003; Frank et al., 2001; Freeman, 1975, 1978; Jansen and Rit, 1995; Jirsa and Haken, 1996; Lopes da Silva et al., 1974; Lopes da Silva et al., 1976; Nunez, 1974; Robinson, 2005; Robinson et al., 1997; Robinson et al., 2001; Rodrigues, 2006; Steyn-Ross et al., 1999; Valdes et al., 1999; Wilson and Cowan, 1972; Wright and Liley, 1996). Models of neuronal dynamics allow one to ask mechanistic questions about how observed data are generated. These questions or

hypotheses can be addressed through model selection by comparing the evidence for different models, given the same data. This endeavour is referred to as DCM (David et al., 2006a; David et al., 2006b; Friston, 2002b; Friston, 2003; Kiebel et al., 2006; Penny et al., 2004a). There has been considerable success in modelling fMRI, EEG, MEG and LFP data using DCM (David et al., 2006a; David et al., 2006b; Garrido et al., 2007b; Kiebel et al., 2006; Moran et al., 2007). All these models embed key nonlinearities that characterise real neuronal interactions. The most prevalent models are called neural-mass models and are generally formulated as a convolution of inputs to a neuronal ensemble or population to produce an output. Critically, the outputs of one ensemble serve as input to another, after some static transformation. Usually, the convolution operator is linear, whereas the transformation of outputs (*e.g.*, mean depolarisation of pyramidal cells) to inputs (firing rates in pre-synaptic inputs) is a nonlinear sigmoidal function. This function generates the nonlinear behaviours that are critical for modelling and understanding neuronal activity. We will refer to these functions as activation or input-firing curves.

The mechanisms that cause a neuron to fire are complex (Destexhe and Pare, 1999; Mainen and Sejnowski, 1995); they depend on the state (open, closed; active, inactive) of several kinds of ion channels in the postsynaptic membrane. The configuration of these channels depends on many factors, such as the history of presynaptic inputs and the presence of certain neuromodulators. As a result, neuronal firing is often treated as a stochastic process. Random fluctuations in neuronal firing function are an important aspect of neuronal dynamics and have been the subject of much study. For example, (Miller and Wang, 2006) looked at the temporal fluctuations in firing patterns in working memory models with persistent states. One perspective on this variability is that it is caused by *fluctuations in the threshold* of the input-firing curve of individual neurons. This is one motivation for a sigmoid activation function at the level of population dynamics; which rests on the well-known result that the average of many different threshold functions is a nonlinear sigmoid. An alternative point of view is that the variability could be caused by thermal noise due to the passive membrane resistance which also has a Gaussian distribution for large ensembles (Manwani and Koch, 1999). We will show the same sigmoid function can be motivated by assuming *fluctuations in the neuronal states*

(Hodgkin and Huxley, 1952). This is a more plausible assumption because variations in postsynaptic depolarisation over a population are greater than variations in firing threshold (Fricker et al., 1999): in active cells, membrane potential values fluctuate by up to about 20mV, due largely to hyperpolarisations that follow activation. In contrast, firing thresholds vary up to only 8mV. Furthermore, empirical studies show that voltage thresholds, determined from current injection or by elevating extracellular K^+, vary little with the rate of membrane polarization and that the "speed of transition into the inactivated states also appears to contribute to the invariance of threshold for all but the fastest depolarisations" (Fricker et al., 1999). In short, the same mean-field model can be interpreted in terms of random fluctuations on the firing thresholds of different neurons or fluctuations in their states. The latter interpretation is probably more plausible from a neurobiological point of view and endows the sigmoid function parameters with an interesting interpretation, which we exploit in this Chapter. It should be noted that (Wilson and Cowan, 1972) anticipated that the sigmoid could arise from a fixed threshold and population variance in neural states; after Equation 1 of their seminal paper they state: "Alternatively, assume that all cells within a subpopulation have the same threshold, … but let there be a distribution of the number of afferent synapses per cell.". This distribution induces variability in the afferent activity seen by any cell.

This is the first in a series of Chapters that addresses the importance of high-order statistics (*i.e.*, variance) in neuronal dynamics, when trying to model and understand observed neurophysiological time-series. In this Chapter, we focus on the origin of the sigmoid activation function, which is a ubiquitous component of many neural-mass and neural-field models. In brief, this treatment provides an interpretation of the sigmoid function as the cumulative density on post-synaptic depolarisation over an ensemble or population of neurons. Using real EEG data we will show that population variance, in the depolarisation of neurons in somatosensory sources generating sensory evoked potentials (SEP) (Litvak et al., 2007) can be quite substantial, especially in relation to evoked changes in the mean. In a subsequent Chapter, we will present a mean-field model of population dynamics that covers both the mean and variance of neuronal states. A special case of this model is the neural-mass model, which assumes the variance is fixed (David et al., 2006a; David et al., 2006b; Kiebel

et al., 2006). In a final Chapter, we will use these models as probabilistic generative models (*i.e.*, dynamic causal models) to show that population variance can be an important quantity, when explaining observed EEG and MEG responses.

This Chapter comprises three sections. In the first, we present the background and motivation for using sigmoid activation functions. These functions map mean depolarisation, within a neuronal population, to expected firing rate. We will illustrate the origins of their sigmoid form using a simple conductance-based model of a single population. We rehearse the well-known fact that threshold or Heaviside operators in the equations of motion for a single neuron lead to sigmoid activation functions, when the model is formulated in terms of mean neuronal states. We will show that the sigmoid function can be interpreted as the cumulative density function on depolarisation, within a population.

In the second section we emphasise the importance of variance or dispersion by noting that a change in variance leads to a change in the form of the sigmoid function. This changes the transfer function of the system and its input-output properties. We will illustrate this by looking at the Volterra kernels of the model and computing the modulation transfer function to show how the frequency response of a neuronal ensemble depends on population variance.

In the final section, we estimate the form of the sigmoid function using the established dynamic causal modelling technique and SEPs, following medium nerve stimulation. In this analysis, we focus on a simple DCM of brainstem (BS) and somatosensory sources, each comprising three neuronal populations. Using standard variational techniques, we invert the model to estimate the density on various parameters, including the parameters controlling the shape of the sigmoid function. This enables us to estimate the implicit probability density function on depolarisation of neurons within each population. We conclude by discussing the implications of our results for neural-mass models, which ignore the effects of population variance on the evolution of mean activity. We use these conclusions to motivate a more general model of population dynamics that will be presented in the subsequent Chapter 5.

4.2 Theory

In this section, we will show that the sigmoid activation function used in neural-mass models can be derived from straightforward considerations about single-neuron dynamics. To do this, we look at the relationship between variance introduced at the level of individual neurons and their population behaviour.

Saturating nonlinear activation functions can be motivated by considering neurons as binary units; *i.e.*, as being in an active or inactive state. (Wilson and Cowan, 1972) showed that (assuming neuronal responses rest on a threshold or Heaviside function of activity) any unimodal distribution of thresholds results in a sigmoid activation function at the population level. This can be seen easily by assuming a distribution of thresholds within a population characterized by the density, $p(w)$. For unimodal $p(w)$, the response function, which is the integral of the threshold density, will have a sigmoid form. For symmetric and unimodal distributions, the sigmoid is symmetric and monotonically increasing; for asymmetric distributions, the sigmoid loses point symmetry around the inflection point; in the case of multimodal distributions, the sigmoid becomes wiggly (monotonically increasing but with more than one inflexion point). Another motivation for saturating activation functions considers the firing rate of a neuron and assumes that its time average equals the population average (*i.e.*, activity is ergodic). The firing rate of neurons always shows saturation and hence sigmoid-like behaviour.

Neurons exhibit an outstanding variety of morphological and physiological properties. The dynamical traits of neuron types have been extensively characterized. Close to threshold, there are two distinct types of input-firing curves: type I and type II: the former curves are continuous and represent an increasing analytic function of input. The latter has a discontinuity, where firing starts after some critical input level is reached. These transitions correspond to a bifurcation from equilibrium to a limit-cycle attractor[5]. The type of bifurcation determines the fundamental computational properties of neurons. Type I and II neuronal behaviour can be generated by the same

[5] In all cases, type I cells experience a saddle-node bifurcation on the invariant circle, at threshold. Type II neurons, may have three different bifurcations; i.e., a subcritical Hopf bifurcation (most frequent), a supercritical Hopf bifurcation, or a saddle node bifurcation outside the invariant circle.

neuronal model (Izhikevich, 2007). From these considerations, it is possible to deduce population models (Dayan, 2001).

We will start with the following ordinary differential equation (ODE) modelling the dynamics of a single neuron from the neural-mass model for EEG/MEG (David and Friston, 2003; David et al., 2006a; David et al., 2006b; Garrido et al., 2007a; Kiebel et al., 2006; Moran et al., 2007); for example, the *i*-th neuron in a population of excitatory spiny stellate cells in the granular layer:

$$\begin{aligned}\dot{x}_1^{(i)} &= x_2^{(i)} \\ \dot{x}_2^{(i)} &= \kappa G(\left\langle H(x_1^{(j)} - w^{(j)})\right\rangle_j + Cu) - 2\kappa x_2^{(i)} - \kappa^2 x_1^{(i)}\end{aligned}. \tag{4.1}$$

This is a fairly ubiquitous form for neuronal dynamics in many neural-mass and cortical-field models and describes neuronal dynamics in terms of two states; $\dot{x}_1^{(i)}$ which can be regarded as depolarisation and $\dot{x}_2^{(i)}$, which corresponds to a scaled current. These ordinary differential equations correspond to a convolution of input with a 'synaptic' differential alpha-function (*e.g.*, (Gerstner, 2001)). This synaptic kernel is parameterised by G, controlling the maximum postsynaptic potential and κ, which represents a lumped rate-constant. Here input has exogenous and endogenous components: exogenous input is injected current u scaled by the parameter, C. Endogenous input arise from connections with other neurons in the same population (more generally, any population). It is assumed that each neuron senses all others, so that the endogenous input is the expected firing over neurons in the population. Therefore, neural mass models are necessarily associated with a spatial scale over which the population is deployed; i.e. the so-called mesoscale[6], from a few hundred to a few thousand neurons.

[6] Different descriptions pertain to at least three levels of organization. At the lowest level we have single neurons and synapses (microscale) and at the highest, anatomically distinct brain regions and inter-regional pathways (macroscale). Between these lies the level of neuronal groups or populations (mesoscale) (Sporns, O., Tononi, G., Kotter, R., 2005. The human connectome: A structural description of the human brain. PLoS Comput Biol 1, e42..

The firing of one neuron is assumed to be a Heaviside function[7] of its depolarisation that is parameterised by some neuron-specific threshold, $w^{(i)}$. We can write this in terms of the cumulative density over the states and thresholds of the population

$$\left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_j = \iint H(x_1 - w)\, p(x_1, w) dx_1 dw = \iint_{x_1 > w} p(x_1, w) dx_1 dw . \tag{4.2}$$

This expression can be simplified, if we assume the states have a large variability in relation to the thresholds (see (Fricker et al., 1999)) and replace the density on the thresholds, $p(w)$ with a point mass at its mode, w. Under this assumption, the input from other neurons can be expressed as a function of the sufficient statistics[8] of the population's states; for example, if we assume a Gaussian density $p(x_1) = N(x_1 : \mu_1, \sigma_1^2)$ we can write

$$\begin{aligned} \left\langle H(x_1^{(j)} - w) \right\rangle_j &= \int_w^\infty p(x_1) dx_1 = S(\mu_1 - w) \\ &\Rightarrow p(x_1) = S'(x_1 - \mu_1) \end{aligned} . \tag{4.3}$$

Where $S(\cdot)$ is the sigmoid cumulative density of a zero-mean normal distribution with variance σ_1^2 (*c.f.*, (Freeman, 1975)). Equation 4.3 is quite critical because it links the motion of a single-neuron to the population density and therefore couples microscopic and mesoscopic dynamics. Finally, we can summarise the population dynamics in terms of the sufficient statistics of the states to give a mean-field model $\dot{\mu} = f(\mu, u)$ by taking the expectation of Eq. 4.1

[7] The linearization using a Heaviside function is naturally not appealing if the states populate the nonlinear tails of the sigmoid function.

[8] The quantities that specify a probability density; *e.g.*, the mean and variance.

$$
\begin{aligned}
\dot{x}_1^{(i)} &= x_2^{(i)} \\
\dot{x}_2^{(i)} &= \kappa G(S(\mu_1 - w) + Cu) - 2\kappa x_2^{(i)} - \kappa^2 x_1^{(i)} \Rightarrow \\
& . \\
\dot{\mu}_1 &= \mu_2 \\
\dot{\mu}_2 &= \kappa G(S(\mu_1 - w) + Cu) - 2\kappa\mu_2 - \kappa^2 \mu_1
\end{aligned}
\tag{4.4}
$$

We can do this easily because the equations of motion are linear in the states (note the sigmoid is not a function of the states). The ensuing mean-field model has exactly the same form as the neural-mass model we use in dynamic causal modelling of electromagnetic observations (David et al., 2006a). It basically describes the evolution of mean states that are observed directly or indirectly. In these neural-mass models the sigmoid has a fixed form[9]

$$
S(\mu_i - w) = \frac{1}{1 + \exp(-\rho(\mu_i - w))}. \tag{4.5}
$$

where ρ is a parameter that determines its slope (*c.f.*, voltage-sensitivity). It is this function that endows the model with nonlinear behaviour and biological plausibility. However, this form assumes that the variance of the states is fixed, because the sigmoid encodes the density on neuronal states (see Eq. 4.3). In the particular parameterisation of Eq. 4.5, the slope-parameter corresponds roughly to the inverse variance or precision of $p(x_i)$; more precisely

$$
\begin{aligned}
\sigma_i^2(\rho) &= \int (x_i - \mu_i)^2 p(x_i) dx_i, \\
p(x_i) &= S'(x_i - \mu_i) = \frac{\rho \exp(-\rho(x_i - \mu_i))}{(1 + \exp(-\rho(x_i - \mu_i)))^2}.
\end{aligned}
\tag{4.6}
$$

Figure 4.1 shows the implicit standard deviation over neural states as a function of the slope-parameter, ρ. Heuristically, a high voltage-sensitivity or gain corresponds to a tighter distribution of voltages around the mean, so that near-threshold increases in

[9] By fixed we mean constant over time. Note that we ignore a constant term here that can be absorbed into exogenous input.

the mean cause a greater proportion of neurons to fire and an increased sensitivity to changes in the mean.

This analysis is based on the assumption that variations in threshold are small, in relation to variability in neuronal states themselves. Clearly, in the real brain, threshold variance is not zero; in the *Appendix* E we show that if we allow for variance on the thresholds, the standard deviation in Figure 4.1 becomes an upper bound on the population variability of the states. In the next section, we look at how the dynamics of a population can change profoundly when the inverse variance (*i.e.*, gain) changes.

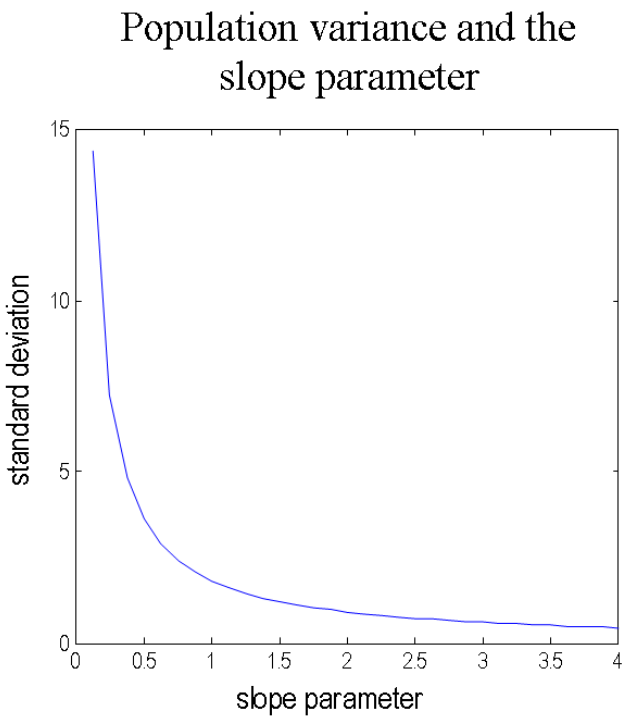


Figure 4.1: Relationship between the sigmoid slope ρ and the population variance, expressed as the standard deviation.

4.3 Kernels, transfer functions and the sigmoid

In this section, we illustrate the effect of changing the slope-parameter (*i.e.*, variance of the underlying neuronal states) on the input-output behaviour of neuronal populations. We will start with a time-domain characterisation, in terms of convolution kernels and conclude with a frequency-domain characterisation, in terms of transfer functions. We will see that the effects of changing the implicit variance are mediated largely by first-order effects and can be quite profound.

4.3.1 Nonlinear analysis and Volterra kernels

The input-output behaviour of population responses can be characterised in terms of a Volterra series. These series are a functional expansion of a population's input that produces its outputs (where the outputs from one population constitute the inputs to another). The existence of this expansion suggests that the history of inputs and the Volterra kernels represent a complete and sufficient specification of population dynamics (Friston et al., 2003). The theory states that, under fairly general conditions, the output y of a nonlinear dynamic system can be expressed in terms of an infinite sum of integral operators

$$y(t) = \sum_i \int \dots \int k_i(\sigma_1, \dots \sigma_i) u(t-\sigma_1) u(t-\sigma_i) d\sigma_1 \dots d\sigma_i \tag{4.7a}$$

where the i-th order kernel is

$$k_i(\sigma_1, \dots \sigma_i) = \frac{\partial y^i(t)}{\partial u(t-\sigma_1) \dots \partial u(t-\sigma_i)} \tag{4.7b}$$

Volterra kernels represent the causal input-output characteristics of a system and can be regarded as generalised impulse response functions (*i.e.*, the response to an impulse or spike). The first-order kernel $\kappa_1(\sigma_1) = \partial y(t) / \partial u(t-\sigma_1)$ encodes the response evoked by a change in input at $t-\sigma_1$. In other words, it is a time-dependent measure of *driving* efficacy. Similarly the second-order kernel

$\kappa_2(\sigma_1,\sigma_2) = \partial y^2(t)/\partial u(t-\sigma_1)\partial u(t-\sigma_2)$ reflects the *modulatory* influence of the input at $t-\sigma_1$ on the response evoked by input at $t-\sigma_2$; and so on for higher orders.

Volterra series have been described as a 'power series with memory' and are generally thought of as a high-order or nonlinear convolution of inputs to provide an output. Essentially, the kernels are a re-parameterisation of the system that encodes the input-output properties directly, in terms of impulse response functions. In what follows, we computed the first and second-order kernels (*i.e.*, impulse response functions) of the neural-mass models, using different slope-parameters. This enabled us to see whether the changes in population variance are expressed primarily in first or second-order effects.

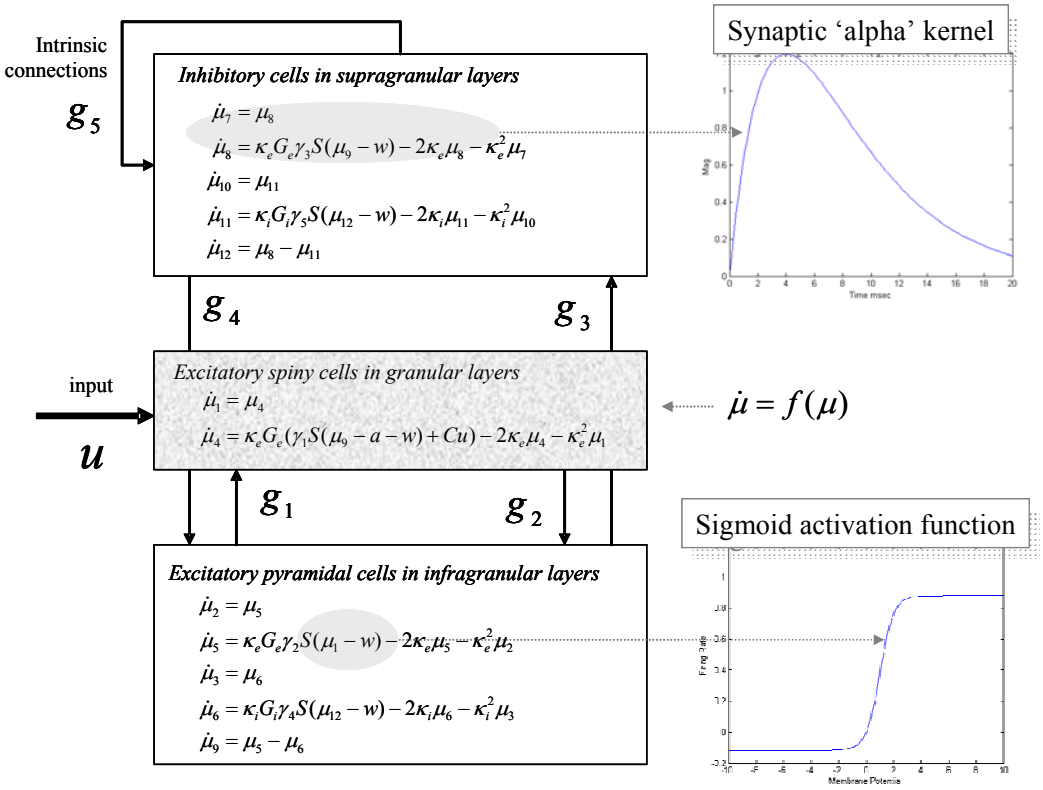


Figure 4.2: Schematic of the neural-mass model used to model a single source (Moran et al., 2007).

The specific neural-mass model we used has been presented in detail by (Moran et al., 2007). This model uses intrinsic coupling parameters, g_i, between three

subpopulations within any source of observed electromagnetic activity. Each source comprises an inhibitory subpopulation in the supragranular layer and excitatory pyramidal (output) population in an infra-granular layer. Both these populations are connected to an excitatory spiny (input) population in the granular layer. This model differs from the model used by ((David and Friston, 2003); see Figure 2.3) in two ways: (i) the inhibitory subpopulation has recurrent self-connections and (ii) spike-rate adaptation is included to mediate slow neuronal dynamics. The equations of motions for a three-population source are shown in Figure 4.2; these all have the form of Eq. 4.4.

Parameter	*Physiological Interpretation*	*Value*
$H_{e,i}$	Maximum Postsynaptic Potentials	*8mV, 32mV*
$\tau_{e,i} = 1/\kappa_{e,i}$	Postsynaptic rate constants	*4ms, 16ms*
$\tau_a = 1/\kappa_a$	Adaptation rate constant	*512 ms*
$\gamma_{1,2,3,4,5}$	Intrinsic connectivity	*128, 128, 64, 64, 4*
w	Threshold	*1.8*

Table 4.1: Model parameters

The first-order kernels or response functions for the depolarisation of the two excitatory populations are shown in Figure 4.3 (upper panels) and the second-order kernels for the excitatory pyramidal cell population are shown in the lower panels, for two values of the slope-parameter; $\rho = 0.8$ and $\rho = 1.6$. The other parameters were chosen such that the system was dynamically stable; see (Moran et al., 2007) and Table 4.1. The kernels were computed as described in the appendix of (Friston et al., 2000).

The first-order responses exhibit a more complicated response for the smaller value of ρ; with pronounced peaks at about 10ms and 20ms for the stellate and pyramidal populations respectively. Both responses resemble damped fast oscillations in the gamma range (about 40Hz). In addition, there appears to be a slower dynamic, with late peaks at about 100ms. This is lost with larger values of ρ (right panels); furthermore, the pyramidal response is attenuated and more heavily damped. The

second-order kernels have two pronounced off-diagonal wing-shaped positivities that do not differ markedly for the two values of ρ. These high-order kernels tell us about nonlinear or modulatory interactions among inputs and speak to asynchronous coupling. For example, the peaks in the second-order kernel at 10ms and 20ms (upper arrow) mean that the response to an input 10ms in the past is positively modulated by an input 20ms ago (and *vice versa*). The long-term memory of the population dynamics is expressed in positive asynchronous interactions (lower arrow) around 100ms. These second-order effects correspond to interactions between inputs at different times, in terms of producing changes in the output. They can be construed as input effects that interact nonlinearly with intrinsic states, which 'remember' the inputs. In the present context, these effects are due to, and only to, the nonlinear form of the sigmoid function, which is mandated by the fact it is a cumulative probability density function. This is an important observation, which means, under the models considered here, population dynamics must necessarily exhibit nonlinear responses.

The effect of changing the gain or slope-parameter is much more evident in the first-order, relative to the second-order kernels. This suggests population variance does not, in itself, change the nonlinear properties of the population dynamics, compared to linear effects. The reason that the slope parameter has quantitatively more marked effects on the first-order kernel is that our neural mass model is only weakly nonlinear; it does not involve any interactions among the states, apart from those mediated by the sigmoid activation function. We can use this to motivate a focus on linear effects using linear systems theory in the frequency domain.

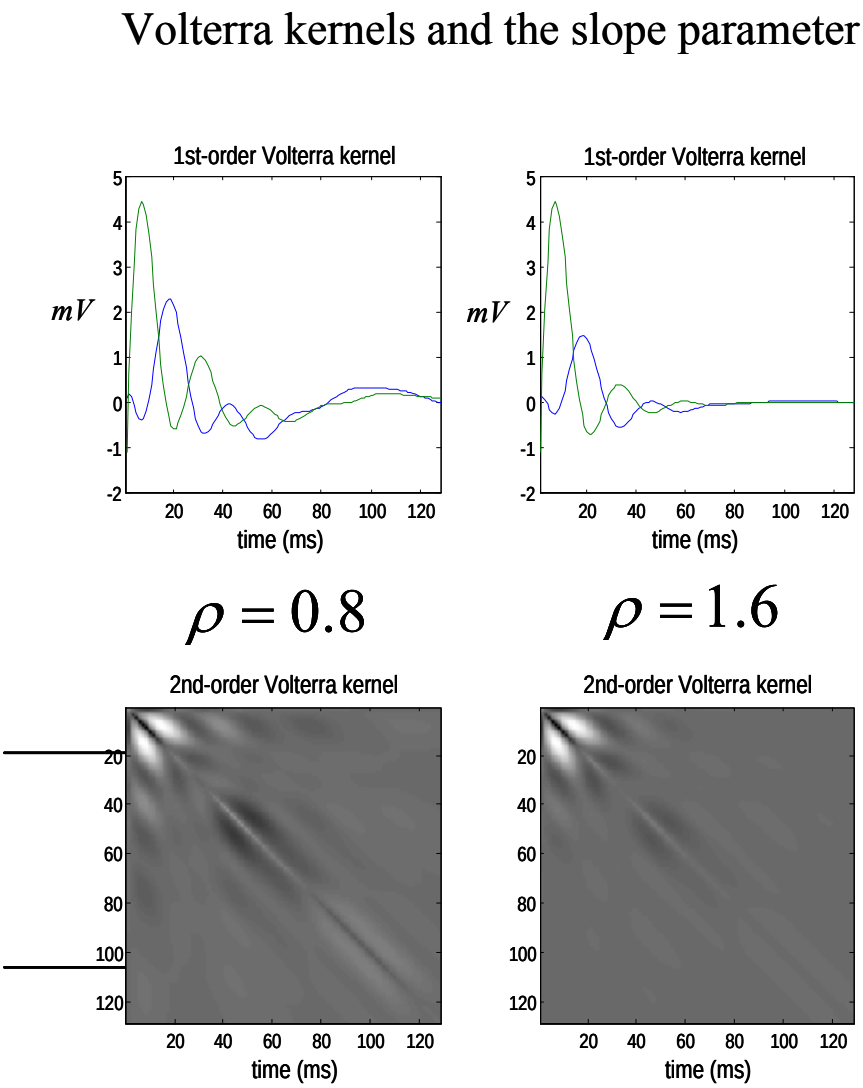


Figure 4.3: Upper Panels: The first-order Volterra kernels for the depolarisation of pyramidal (blue) and spiny stellate (green) populations, for two different values of ρ (left: 0.8, right: 1.6). There is a difference between the waveform, which is marked for the pyramidal cells. **Lower panels**: The corresponding second-order Volterra kernels in image format.

4.3.2 Linear analysis and transfer functions

An alternative characterisation of generalised kernels is in terms of their Fourier transforms, which furnish generalised transfer functions. A transfer function allows one to compute the frequency or spectral response of a population given the spectral

characteristics of its inputs. (Moran et al., 2007) have presented a linear transfer function analysis of this neural-mass model previously. Our model is linearised by replacing the sigmoid function with a first-order expansion around $\mu_i = 0$ to give

$$S(\mu_i) = S'(-w)\mu_i . \tag{4.8}$$

This assumes small perturbations of neuronal states around steady-state. Linearising the model in this way allows us to evaluate the transfer function

$$\mathrm{H}(s) = C(sI - A)^{-1}B \tag{4.9}$$

where the state matrices, $A = \partial f / \partial x$ and $B = \partial f / \partial u$ are simply the derivatives of the equations of motion (*i.e.*, Eq. 4.4) with respect to the states and inputs respectively. The frequency response for steady-state input oscillations at ω radians per second, obtains by evaluating the transfer function at $s = j\omega$ (where $j\omega$ represents the axis of the complex *s*-plane corresponding to steady-state frequency responses). When the system is driven by exogenous input with spectrum, $U(j\omega)$, the output is the frequency profile of the stimulus modulated by the transfer function

$$\left|Y(j\omega)\right| = \left|\mathrm{H}(j\omega)\right|\left|U(j\omega)\right| \tag{4.10}$$

In brief, the transfer function, $H(s)$, filters or shapes the frequency spectra of the input, $U(s)$ to produce the observed spectral response, $Y(s)$. The transfer function $H(s)$ represents a normalized model of the systems input-output properties and embodies the steady-state behaviour of the system. Eq. 4.9 results from one of the most useful properties of the Laplace transform, which enables differentiation to be cast as a multiplication. One benefit of this is that convolution in the time domain can be replaced by multiplication in the *s*-domain. This reduces the computational complexity of the calculations required to analyze the system.

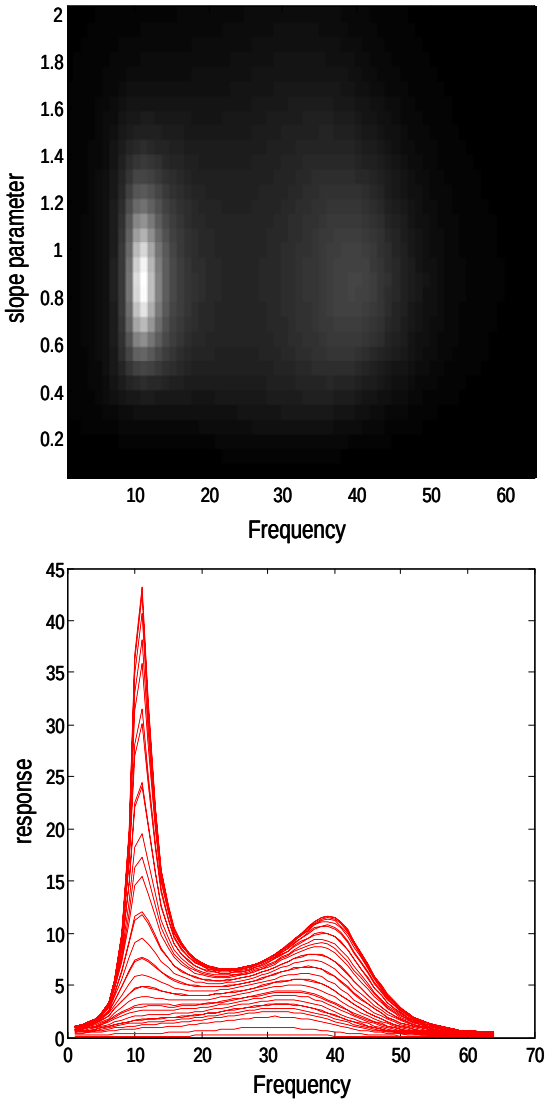


Figure 4.4: Upper Panel: Image of the transfer function magnitude $H(s)$ where ρ is varied from a sixteenth to two. **Lower Panel**: Plot of the same data over frequencies.

We examined the effects of the slope-parameter on the transfer function by computing $|\mathrm{H}(j\omega)|$ for different values of $\rho = \frac{1}{16}, \frac{2}{16}, \ldots, 2$. $|\mathrm{H}(j\omega)|$ corresponds to the spectral response under white noise input (see Eq. 4.10). Figure 4.4 shows the spectral

response is greatest at about, $\rho = 0.8$ when it exhibits a bimodal frequency distribution; with a pronounced alpha peak (~12Hz) and a broader gamma peak (~40Hz). As ρ increases or decreases from this value the alpha component is lost, leading to broad-band responses expressed maximally in the gamma-range. This is an interesting result, which suggests that the population's spectral responses are quite sensitive to changes in the dispersion of states, particularly with respect the relative amount of alpha and gamma power. Having said this, these results should not be generalised because they only hold for the values of the other model parameters we used. These values were chosen to highlight the dependency on the slope-parameter.

To illustrate the change in the response properties caused by a change in ρ, we computed the response of the excitatory populations to an input spike embedded in white noise process (where the amplitude of the noise was one sixteenth of the spike). Using exactly the same input, the responses were integrated for two the values of ρ above: $\rho = 0.8$ which maximises the frequency response and a larger value, $\rho = 1.6$. Figure 4.5 shows the ensuing depolarization of pyramidal and spiny cells and corresponding time-frequency plots. For the smaller ρ (large population variance), the output is relatively enduring with a predominance of alpha power. For the larger value (small population variance), the output is more transient and embraces higher frequencies. We will return to this distinction in an empirical setting in the next section, where we try to estimate the slope-parameters and implicit population variance using real data.

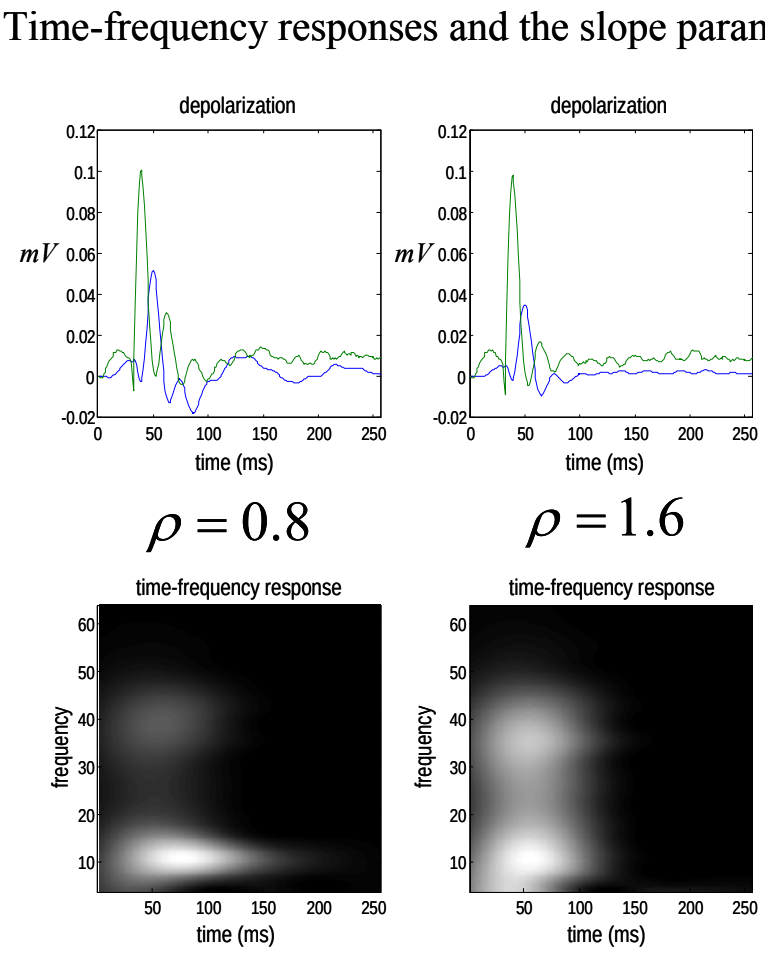


Figure 4.5: Upper Panels: Integrated response to a noisy spike input, for two different values of ρ (left: 0.8, right: 1.6). The response of the excitatory pyramidal (output) population is shown in blue, and the response of the spiny stellate in green. **Lower panels**: the respective time-frequency responses for the two ρ cases.

4.4 Estimating population variance with DCM

In this final section, we exploit the interpretation of the sigmoid as a cumulative density on the states, specifically the depolarisation. This interpretation renders the derivative of the sigmoid a probability density on the voltage: recall from the first section

$$\int_{w}^{\infty} p(x_i)dx_i = S(\mu_i - w) \Rightarrow p(x_i) = S'(x_i - \mu_i) \tag{4.11}$$

This means we can use estimates of the slope-parameter, which specifies S', to infer the underlying variance of depolarisation in neuronal populations (or an upper bound; see *Appendix* E). In what follows, we estimate the slope-parameter using EEG data and Dynamic Causal Modelling. We present two analyses. The first addressed the question: "are changes in the mean depolarisation small or large relative to the dispersion of voltages?" We answered this by evaluating the evoked changes in mean depolarisation in somatosensory sources generating SEPs and then comparing the amplitude of these perturbations with the implicit variance. The second analysis tried to establish whether population variance is stable over time. This issue has profound implications for neural-mass models that assume variance does not change with time.

4.4.1 Analysis of somatosensory responses

We analyzed data from a study of long-term potentiation (LTP) reported in (Litvak et al., 2007). LTP is a long-lasting modification of synaptic efficacy and is believed to represent a physiological substrate of learning and memory (Bliss and Lomo, 1973; Cooke and Bliss, 2006; Malenka and Bear, 2004; Martin et al., 2000). (Litvak et al., 2007) used paired associative stimulation (PAS), which involved repetitive magnetic cortical stimulation timed to interact with median nerve (MN) stimulation-induced peripheral signals from the hand. The PAS paradigm has been shown to induce long-lasting changes in MN somatosensory evoked potentials MN-SSEP; (Wolters et al., 2005) as measured by single-channel recordings from the scalp region overlying somatosensory cortex. The generators of MN-SSEPs evoked by compound nerve stimulation have been studied extensively with both invasive and non-invasive methods in humans and in animal models (for a review see (Allison et al., 1991)). (Litvak et al., 2007) characterised the topographical distribution of PAS-induced excitability changes as a function of the timing and composition of afferent (MN) somatosensory stimulation, with respect to transcranial magnetic stimulation (TMS).

In this work, we analysed the SEP data from one subject, following MN stimulation (*i.e.*, in the absence of magnetic stimulation), with DCM. The network architecture was based on reports in published literature ((Buchner et al., 1995; Litvak et al., 2007; Ravazzani et al., 1995)). We modelled the somatosensory system with four equivalent current dipoles or sources, each comprising three neuronal subpopulations as described in the previous section. Exogenous input was modelled with a gamma function (with free parameters), peaking shortly after MN stimulation. In this model, exogenous input was delivered to the brainstem source (BS), which accounts for early responses in the medulla. In Brodmann area (BA) 3b of S1, we deployed three sources, given previous work showing distinct tangential and radial dipoles. We employed a third source to account for any other activity. These sources received endogenous input from the BS source, via extrinsic connections to the stellate cells.

We inverted the resulting DCM using a variational scheme (Friston et al., 2007) and scalp data from 12ms to 100ms, following MN stimulation. This inversion used standard variational techniques, which rest on a Bayesian expectation maximization (**EM**) algorithm under a Laplace approximation to the true posterior (Appendix B). This provided the posterior densities of the models parameters, which included the synaptic parameters of each population, the extrinsic connection strengths, the parameters of the gamma input function and the spatial parameters of the dipoles (for details see (David et al., 2006a; David et al., 2006b; Kiebel et al., 2006)). The resulting posterior means of dipole locations and moments are shown in Figure 4.6 (upper panel).

In terms of the temporal pattern of responses, the MN-SSEP has been studied extensively (Allison et al., 1991). A P14 component is generated subcortically, then a N20–P30 complex at the sensorimotor cortex (BA 3b) exhibits a typical 'tangential source pattern'. This is followed by a P25–N35 complex with a 'radial source pattern'. The remainder of the SEP can be explained by an 'orthogonal source pattern' originating from the hand representation in S1 (Litvak et al., 2007). These successive response components were reproduced precisely by the DCM. The accuracy of the DCM can be appreciated by comparing the observed data with predicted responses in Figure 4.6 (lower panels).

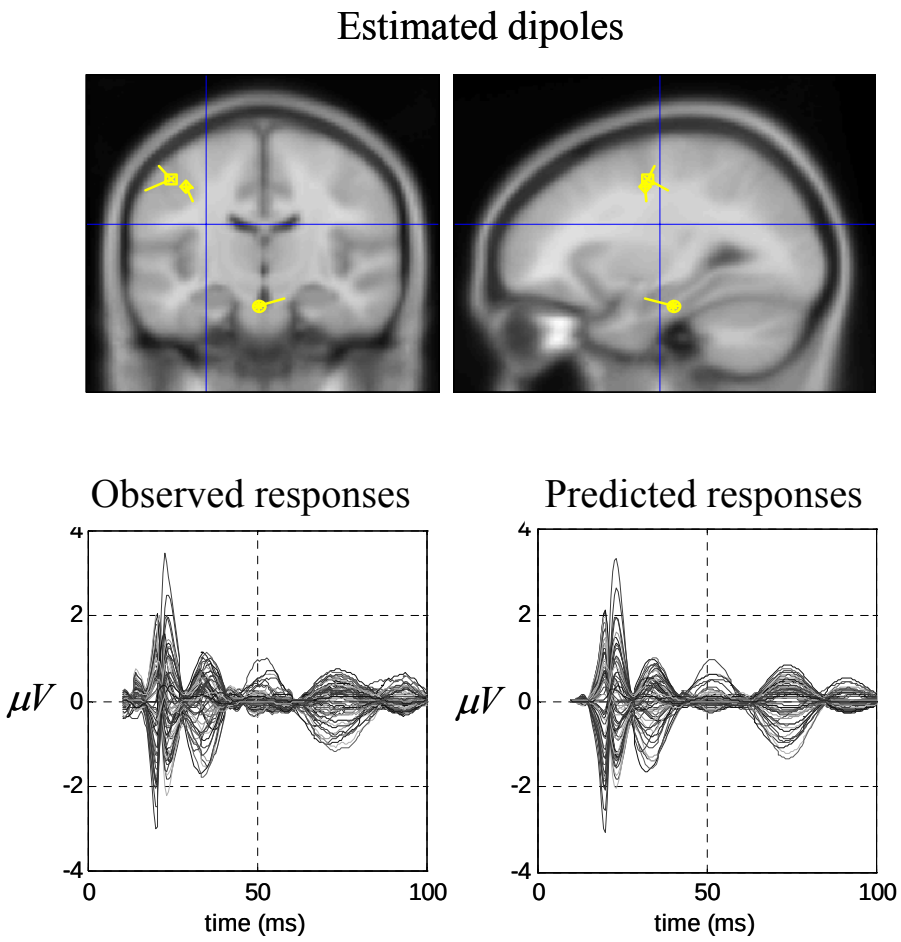


Figure 4.6: Upper Panels: Source locations estimated with DCM: Orthogonal slices showing the brainstem dipole (BS) and the left primary somatosensory cortex (S1) source (consisting of three dipoles: tangential, radial and orthogonal). **Lower panels**: The left graph shows the observed MN-SSEP in channel space. The right graph demonstrates the goodness of fit of the DCM using the same format.

Using Eq. 4.6 and the maximum *a posteriori* estimate of the slope-parameter, we evaluated the implicit variance of depolarization $\sigma_i^2(\rho)$ within each neuronal population (see Equation 4.6 and Figure 4.1). This variance can be combined with the time-dependent mean depolarisation $\mu_i(t)$ of any population, estimated by the DCM, to reconstruct the implicit density on population depolarisation over peristimulus time. Figure 4.7 shows this density in terms of its mean and 90% confidence intervals for the first S1 pyramidal population. This quantitative analysis is quite revealing; it

shows that evoked changes in the mean depolarisation are small in relation to the dispersion. This means that only a small proportion of neurons are driven above threshold, even during peak responses. For example, using the estimated threshold, *w*, during peak responses only about 12% of neurons would be above threshold and contribute to the output of the population. In short, this sort of result suggests that communication among different populations is mediated by a relatively small faction of available neurons and that small changes in mean depolarisation are sufficient to cause large changes in firing rates, because depolarisation is dispersed over large ranges.

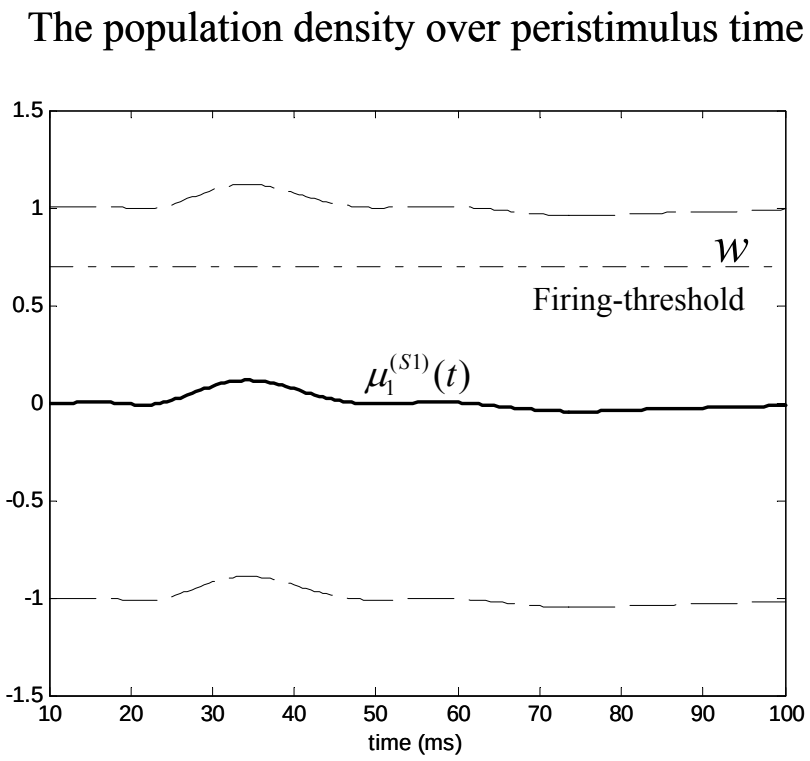


Figure 4.7: S1 source (pyramidal population) mean depolarization (solid line) as estimated by DCM. The variance is depicted with 90% confidence intervals (dashed lines); *i.e.*, $\pm 1.641 \times \sigma_i^2(\rho)$.

4.5 Epilogue

The preceding analysis assumes that the variance is fixed over peristimulus time. Indeed neural-mass models in general assume a fixed variance because they assume a fixed-form for the sigmoid activation function. Neural-mass models are obliged to make this assumption because their state variables allow only changes in mean states, not changes in variance or higher-order statistics of neuronal activity. The question is: Is this assumption sensible?

To answer this question, in the next chapter, mean-field models are compared that cover both the mean and variance as time-varying quantities. Under the neural-mass model considered here, one cannot test formally for changes in variance. However, one can provide anecdotal evidence for changes in variance by estimating the slope-parameters over different time-windows of the data. If the variance does not change with time, then the estimate of population variance should not change with the time-window used to estimate it. Figure 4.8 show estimates of ρ (with 90% confidence intervals)[10] that obtain using different time-windows of the MN-SSEP data. For example, the estimate, ρ_{80} was obtained using the time period from 10 to 80ms. It can be seen immediately that the slope-parameter and implicit variance changes markedly with the time-window analysed.

However, the results in Figure 4.8 should not be over-interpreted because there are many factors that can lead to differences in the conditional density when the data change; not least a differential shrinkage to the prior expectation. However, this instability in the conditional estimates speaks to the potential importance of modelling population variance as a dynamic quantity.

[10] Note that these confidence intervals are not symmetric about the mean. This is because we actually estimate $\ln \rho$, under Gaussian shrinkage priors. Under the Laplace assumption (Friston, et al 2007) this means the condition density $q(\rho)$ has a log-normal form.

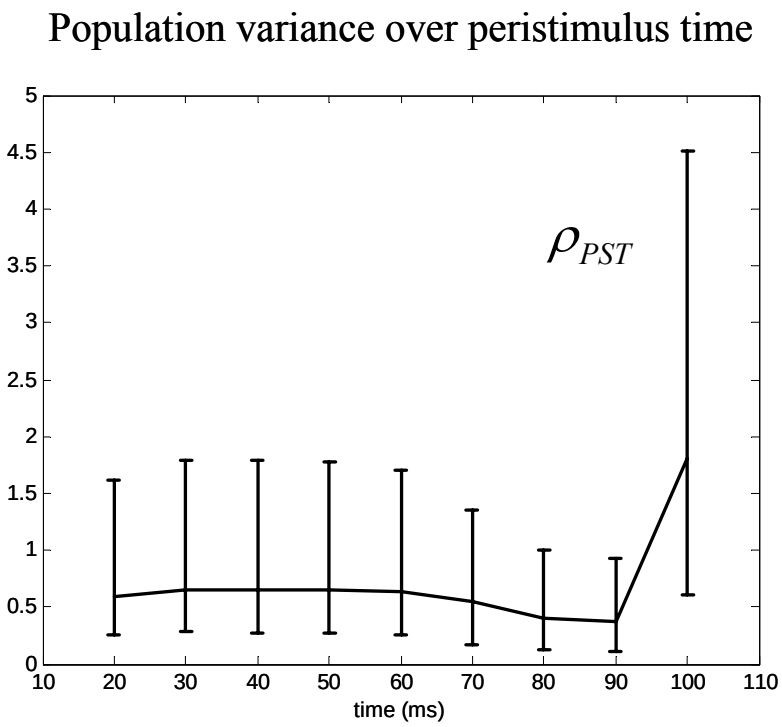


Figure 4.8: Change in the conditional estimates of ρ (mean and 90% confidence intervals) as a function of the peri-stimulus time-window used for model inversion.

We have seen that the dynamics of neuronal populations can be captured qualitatively via a system of coupled differential equations, which describe the evolution of the average firing rate of each population. To accommodate stochastic models of neural activity, one could solve the associated Fokker-Planck equation for the probability distribution of activities in the different neuronal populations. This can be a difficult computational task, in the context of a large number of states and populations (e.g., (Harrison et al., 2005)). (Rodriguez and Tuckwell, 1998; Rodriguez, 1998) presented an alternative approach for noisy systems using the method of moments (MM). This entails the derivation of deterministic ordinary differential equations (ODE) for the first and second-order moments of the population density. The resulting reduced system lends itself to both analytical and numerical solution, as compared with the original Langevin formulation.

(Hasegawa, 2003b) proposed a semi-analytical mean-field approximation, in which equations of motions for moments were derived for a FitzHugh-Nagumo (FN)

ensemble. In (Hasegawa, 2003b), the original stochastic differential equations were replaced by deterministic ODEs by applying the method of moments (Rodriguez and Tuckwell, 1998). This approach was applied to an ensemble of Hodgkin-Huxley (HH) neurons, for which effects of noise, coupling strength, and ensemble size have been investigated. In (Deco and Marti, 2007), the MM was extended to cover bimodal densities on the state variables; such that a reduced system of deterministic ODEs could be derived to characterise regimes of multistability. We will use MM in our next Chapter, where we derive the ODEs of the sufficient statistics of integrate-and-fire ensembles of distributed neuronal sources. These ODEs will form the basis of dynamical causal models of empirical data in the last Chapter.

4.6 Conclusion

In this Chapter, our focus was on how the sigmoid activation function, linking mean population depolarization to expected firing rate, can be understood in terms of the variance or dispersion of neuronal states. We showed that the slope-parameter ρ models formally the effects of variance (to a first approximation) on neuronal interactions. Specifically, we saw that the sigmoid function can be interpreted as a cumulative density function on depolarisation, within a population. Then, we looked at how the dynamics of a population can change profoundly when the variance (slope-parameter) changes. In particular, we examined how the input-output properties of populations depend on ρ, in terms of first (driving) and second (modulatory) order convolution kernels and corresponding transfer functions.

We used real EEG data to show that population variance, in the depolarisation of neurons from somatosensory sources generating SEPs, can be quite substantial. Using DCM, we estimated the SEP parameter density controlling the shape of the sigmoid function. This allowed us to quantify the population variance in relation to the evolution of mean activity of neural-masses. The quantitative results of this analysis suggested that only a small proportion of neurons are actually firing at any time, even during the peak of evoked responses.

The insights from these studies motivate a more general model of population dynamics that will be presented in the next Chapters; where we will compare DCMs based on density-dynamics with those based on neural-mass models. Modelling the interactions between mean neuronal states (*e.g.*, depolarisation) and their dispersion or variance over each population may provide a better and more principled model of real data. In brief, these models allow us to ask if the variance of neuronal states in a population affects the mean (or *vice versa*) using the evidence or marginal likelihood of the data under different models. Moreover, we can see if observed responses are best explained by mean firing rates, or some mixture of the mean and higher-order moments. This will allow one to adjudicate between models that include high-order statistics of neuronal states in EEG time-series models. In a final Chapter, we will use these models as probabilistic generative models (*i.e.*, dynamic causal models) to show that population variance is an important quantity, when explaining observed EEG and MEG responses.

CHAPTER 5

POPULATION DYNAMICS UNDER THE LAPLACE ASSUMPTION

In the previous Chapter, we saw how the sigmoid activation function, linking mean population depolarization to expected firing rate can be understood in terms of the variance or dispersion of neuronal states. We saw that the sigmoid function can be interpreted as a cumulative density function on depolarisation, within a population. This motivates a more general model of population dynamics that will be presented here. In this Chapter, we describe a generic approach to modelling dynamics in neuronal populations. This approach models a full density on the states of neuronal populations but finesses this high-dimensional problem by re-formulating density dynamics in terms of ordinary differential equations on the sufficient statistics of the densities considered (c.f., the method of moments). The particular form for the population density we adopt is a Gaussian density (c.f., the Laplace assumption). This means population dynamics are described by equations governing the evolution of the population's mean and covariance. We derive these equations from the Fokker-Planck formalism and illustrate their application to a conductance-based model of neuronal exchanges. One interesting aspect of this formulation is that we can uncouple the mean and covariance to furnish a neural-mass model, which rests only on the populations mean. This enables us to compare equivalent mean-field and neural-mass models of the same populations and evaluate, quantitatively, the contribution of population variance to the expected dynamics.

5.1 Introduction

Mean-field models of neuronal dynamics have a long history, spanning a half-century (*e.g.*, (Beurle, 1956)). Models are essential for neuroscience, in the sense that most interesting questions about the brain pertain to neuronal mechanisms and processes that are not directly observable (Breakspear et al., 2006; Tass, 2003). This means that

questions about neuronal function are generally addressed by inference on models or their parameters; where the model links hidden neuronal processes to our observations and questions (Valdes et al., 1999). Broadly speaking, models are used to generate data to study emergent behaviours. Alternatively, they can be used as forward or observation models (e.g., dynamic causal models), which are inverted given empirical data (David et al., 2006a; Kiebel et al., 2006). This inversion allows one to select the best model (*i.e.*, hypothesis), given some data and make probabilistic statements about the parameters of that model (e.g., (Penny et al., 2004a)).

In particular, mean-field models are appropriate for data that reflect the behaviour of neuronal populations, such as the electroencephalogram (EEG), magnetoencephalogram (MEG) and functional magnetic resonance imaging (fMRI) data. The most prevalent models of neuronal populations or ensembles are based upon the so-called *mean-field approximation*. This approximation replaces the time-averaged discharge rate of individual neurons with a common time-dependent population activity (ensemble average; (Haskell et al., 2001; Knight, 2000)). This assumes ergodicity for all neurons in the population. The mean-field approximation is used extensively in statistical physics for otherwise computationally or analytically intractable problems. An exemplary approach, owing to Boltzmann and Maxwell, is the approximation of the motion of molecules in a gas by mean-field terms such as temperature and pressure. Similarly, evoked response potentials (ERPs) represent the average response over millions of neurons, where the mean-field approximation describes the time-dependent distribution of the average population response. This is possible because the dynamics of the mean of the density are much less stochastic than the response of a single neuron. This makes it feasible to develop algorithms that use Bayesian inference to infer neuronal parameters given measured responses, using mean-field models (e.g., (Harrison et al., 2005)).

Usually, neural-mass models are used to model the evolution of the mean response or the response at steady state. Generally, mean-field approximations can be used to model the full distribution of the population response. However, mean-field models can be computationally expensive, because one has to consider the density at all points in neuronal state-space as opposed to a single quantity (e.g., the mean). In this

Chapter, we present an approach that simplifies the mean-field model by using the Laplace approximation: Under the Laplace approximation, the population or ensemble density assumes a Gaussian form, whose sufficient statistics comprise the conditional mean and covariance. In contrast to neural-mass models, this allows one to model interactions between the first two moments (i.e., mean and variance) of neuronal states. In the next Chapter, we will use the Laplace and neural mass approximations presented here as generative models of electrophysiological responses to sensory input. We will use Bayesian model comparison to compare both models and establish whether empirical responses contain evidence for a role of variance in shaping population dynamics. Here, we focus on the models themselves.

The Laplace approximation is a ubiquitous device in statistical physics and machine learning and finesses difficult integration problems when integrating over probability densities (see (Chumbley et al., 2007; Friston et al., 2007)). Exactly the same device is used here to furnish a simple scheme for modelling density dynamics. Because the sufficient statistics of a Gaussian density can be specified in terms of the first two moments, the ensuing scheme is formally identical to the second-moment method described by (Rodriguez, 1996). The method of moments (Rodriguez and Tuckwell, 1998; Rodriguez, 1996), replaces a system of stochastic differential equations (describing the states of an ensemble) with deterministic equations describing the evolution of the sufficient statistics or moments of an ensemble density. This approach was first applied to a FitzHugh-Nagumo (FN) neuron (Rodriguez, 1996; Tuckwell and Rodriguez, 1998) and later to Hodgkin-Huxley (HH) neurons (Rodriguez and Tuckwell, 1998, 2000). This approach assumes that the distributions of the variables are approximately Gaussian so that they can be characterized by their first and second order moments; i.e., the means and covariances. In related work, Hasegawa described a dynamical mean-field approximation (DMA) to simulate the activities of a neuronal network. This method allows for qualitative or semi-quantitative inference on the properties of ensembles or clusters of FN and HH neurons; see (Hasegawa, 2003a; Hasegawa, 2003b).

This Chapter comprises three sections. In the first, we provide the background to modelling neuronal dynamics with mean-field and neural-mass models. In the second

section, we derive a generic mean-field treatment of neuronal dynamics starting with any equations of motion. This treatment is based on a Laplace approximation to the ensemble density, and is formulated compactly, in terms of the equations of motion for the sufficient statistics of the ensemble density. This approach reduces to a neural-mass model when the second-order statistics (*i.e.*, variance) of neuronal states are ignored. We will illustrate how neuronal state equations are reformulated as a mean-field approximation, using a simple conductance-based model (c.f., (Morris and Lecar, 1981)). In the third section, we establish the validity of the Laplace approximation by comparing the response of simulated ensembles of neurons to responses under the Laplace and neural-mass assumptions. The key behaviour we are interested in is the coupling between the mean and variance of the ensemble, which is lost in the neural-mass approximations.

5.2 Mean field and neural-masses

What follows is a brief summary of the material in (Deco et al., 2008)[11], which provides a full account of mean-field models in neuroscience. The most prevalent models of neuronal populations or ensembles are based on the mean-field approximation. The basic idea behind these models is to approximate a very high dimensional probability distribution with the product of a number of simpler (marginal) densities. Its utility is best seen in the context of ensemble or population density models.

5.2.1 Mean-field models

Ensemble models attempt to model the dynamics of large populations of neurons. Any single neuron can have a number of attributes; for example, post-synaptic membrane depolarisation, V, capacitive current I or the time since the last action potential, T. Each attribute induces a dimension in the state or *phase-space* of a neuron. In this example, the phase-space would be three-dimensional and the state of each neuron

[11] There seems to be some inconsistent terminology across different scientific authors from different backgrounds and decades. The terminology used in this piece of work follows this review.

would correspond to a point $x = \{V, I, T\} \in \Re^3$ or particle in phase-space. Imagine a very large number of such neurons that populate phase-space with a density, $q(x,t)$. As the state of each neuron evolves, the points will flow through phase-space and the *ensemble density*; $q(x,t)$ will evolve until it reaches some steady-state or equilibrium. It is the evolution of the density *per se* that is characterised in ensemble density methods. These models are particularly attractive because the density dynamics conform to a simple equation; *the Fokker-Planck equation* (Dayan, 2001; Frank et al., 2001; Gerstner, 2002; Risken, 1996)

$$\begin{aligned}\dot{q} &= -\nabla \cdot fq + \nabla \cdot D\nabla q \\ &= -\sum_{i=1}^{n} \frac{\partial (f_i q)}{\partial x_i} + \sum_{i,j=1}^{n} \left(\frac{\partial}{\partial x_i} D_{ij} \frac{\partial}{\partial x_j} \right) q\end{aligned} \quad (5.1)$$

For *n* states; $x \in \Re^n$. The equation comprises flow and dispersion terms, which embed the assumptions about the dynamics and random fluctuations. The flow, $f(x,t)$ and dispersion, $D(x,t)$ constitute our model at the neuronal level. This level of description is usually framed as a *stochastic differential equation* (SDE) that describes how the states evolve as functions of each other and some random fluctuations

$$dx = f(x)dt + \sigma dw. \quad (5.2)$$

Where, $D = \frac{1}{2}\sigma^2$ and $w(t)$ is a standard *Wiener process* (where, in one dimension $w(t + \Delta t) - w(t) \sim N(0, \Delta t)$). Under the Fokker-Planck formalism, even if the dynamics of each neuron are very complicated, or indeed chaotic, the density dynamics remain simple, linear and deterministic. In short, for any model of neuronal dynamics, specified as a stochastic differential equation, there is a deterministic linear equation that can be integrated to generate ensemble dynamics. However, there is a problem; the dimensionality of phase-space can become unmanageably large, if we consider too many neuronal states or different types of neuron. Generally speaking, full ensemble models of realistic systems are computationally intractable. However, we can use a mean-field approximation to finesse this problem.

5.2.2 The mean-field approximation

Consider the states of *m* sorts of neuron, each with *n* states; then the states $x = x^{(1)},\ldots,x^{(m)} \in \Re^{n\times m}$ could have a large $n \times m$ dimensionality. However, if we assume the density factorises over the *m* populations

$$q(x) \approx \prod_{i=1}^{m} q(x^{(i)}) \tag{5.3}$$

we have only to deal with *n*-dimensional states $x^{(i)} \in \Re^n$. However, by factorising the density into marginal densities we have effectively assumed that they are independent. This implausible assumption can be circumvented by coupling the ensembles so that the flow in the phase-space of one ensemble, $f(x^{(i)},\mu)$ depends upon the others; through mean-field quantities $\mu^{(j)} = \mu(q(x^{(j)}))$. These are *phase-functions* of the ensemble densities. These mean-field effects could come from the same ensemble and model interactions among neurons in the same population. The ensuing dynamics conform to a series of coupled nonlinear Fokker-Planck equations (Frank, 2004). Typically, these phase-functions return the average state (*e.g.*, mean depolarisation or firing). It is important to realise that coupling ensembles through mean-field quantities, $\mu \in \Re^n$ entails strong assumptions about the nature of the interactions: specifically, the dynamics or fluctuations in one member of an ensemble cannot affect a member of another ensemble. Instead, all the neurons in one ensemble are affected identically by the average behaviour of another ensemble. In many instances, this is a reasonable approximation but, clearly, it makes the exact form of the mean-field approximation an important consideration. In the next Chapter, we will incorporate the mean-field model of this Chapter into dynamic causal models of distributed neuronal sources. In this context, the coupling above determines how one neuronal source influences another; i.e., it corresponds to effective connectivity (David et al., 2006a; Friston et al., 2003).

Even for a single ensemble the dimensionality of $x \in \Re^n$ may preclude numerical or analytic analysis. One can simplify the model by summarising the dynamics with a small number of states. In the limit, one can reduce the dynamics to a single neuronal state $x \in \Re$. An important example is when the state is voltage, *i.e.*, $x = \{V\}$. For example, (Gerstner, 2002) formulate the dynamics of an ensemble of *leaky integrate and fire* neurons with equations of motion

$$f(V) = -\frac{g_L}{C}(V - V_L) + \mu \tag{5.4}$$

using the Fokker-Planck equation (Eq. 5.1), with boundary conditions on $q(V,t)$ that model spiking and a re-setting of the membrane potential. Here, C represents membrane capacitance and g_L a leakage conductance An alternative method is to use the auxiliary variable T (time elapsed since last spike) to parameterize the *refractory density,* $q(T,t)$; see (Eggert and van Hemmen, 2001). (Chizhov et al., 2006) have refined this approach to account for fast and slow ionic currents, with some compelling results.

In summary, one can approximate an ensemble density on a high-dimensional phase-space with a series of low-dimension ensembles that are coupled through mean-field effects. The product of these marginal densities is then used to approximate the full density. Critically, the mean-field coupling induces nonlinear dependencies among the density dynamics of each ensemble. This typically requires a nonlinear Fokker-Planck equation for each ensemble. The Fokker-Planck equation prescribes the evolution of the ensemble dynamics, given any initial conditions and equations of motion that constitute our neuronal model. However, it does not specify how to encode or parameterize the density. There are several approaches to density parameterization (Casti et al., 2002; Haskell et al., 2001; Knight, 2000; Nykamp and Tranchina, 2000; Omurtag et al., 2000; Sirovich, 2003). These include binning the phase-space and using a discrete approximation to a free-form density. However, this can lead to a vast number of differential equations, especially if there are multiple states for each population. One solution to this is to reduce the dimension of the phase-space to render the integration of the Fokker-Planck more tractable (e.g., (Chizhov and

Graham, 2007)). Alternatively, one can assume the density has a fixed parametric form and deal only with its sufficient statistics (Hasegawa, 2003a; Hasegawa, 2003b; Rodriguez and Tuckwell, 1998; Rodriguez, 1996). The simplest form is a delta-function or point mass; under this assumption we get neural-mass models.

5.2.3 Neural-mass models

Neural-mass models can be regarded as a special case of ensemble density models, where we summarize the ensemble density with a single number. Early examples can be found in the work of (Beurle, 1956) and (Griffith, 1963, 1965). The term *mass action model* was coined by (Freeman, 1975) as an alternative to density dynamics. Assuming that the equilibrium density has a *point mass* (i.e., a delta function), we can motivate the description above in terms of the expected value of the state, μ; under the assumption that the equilibrium density has a *point mass* (i.e., a delta function). This is one perspective on why these simple mean-field models are called *neural-mass models*. In short, we replace the full ensemble density with a mass at a particular point and then summarize the density dynamics by the location of that mass. What we are left with is a set of non-linear differential equations describing the dynamic evolution of this mode. But what have we thrown away? In the full nonlinear Fokker-Planck formulation, different phase-functions or probability density moments could couple to each other; both within and between ensembles. For example, the average depolarisation in one ensemble could be affected by the dispersion or variance of depolarisation in another, see (Deco et al., 2008). In neural-mass models, one ignores this potential dependency because only the expectations or first moments are coupled. There are several devices that are used to compensate for this simplification. Perhaps the most ubiquitous is the use of a sigmoid function $\varsigma(V)$ relating expected depolarisation to expected firing-rate (Freeman, 1975; Marreiros et al., 2008a). This implicitly encodes variability in the post-synaptic depolarisation, relative to the potential at which the neuron would fire. A common form for neural-mass equations of motion posits a second-order differential equation for expected voltage μ_V or, equivalently, two coupled first-order equations, where

$$\left(\frac{\partial^2}{\partial t^2} + 2\kappa\frac{\partial}{\partial t} + \kappa^2\right)\mu_V = \kappa^2\gamma\varsigma(\mu_V) \Rightarrow$$
$$\dot{\mu}_V = \mu_I \qquad (5.5)$$
$$\dot{\mu}_I = \kappa^2\gamma\varsigma(\mu_V) - 2\kappa\mu_I - \kappa^2\mu_V$$

where μ_I can be regarded as capacitive current. The input $\gamma\varsigma(\mu_V)$ is commonly construed as firing-rate (or pulse-density) and is a sigmoid function of mean voltage of the same of another ensemble. The coupling constant γ scales the amplitude of this mean-field effect. The constant κ controls the rise and decay of the implicit (synaptic) impulse response $G(t)$ to input; convolving input with this impulse response kernel gives the expected voltage

$$\mu_V(t) = \int G(t-t')\varsigma(\mu_V(t))dt' \qquad (5.6)$$
$$G(t) = \gamma\kappa^2 t\exp(-\kappa t)$$

This form of neural-mass model has been used extensively to model electrophysiological recordings (*e.g.*, (David and Friston, 2003; Elbert et al., 1994; Jansen and Rit, 1995; Kincses et al., 1999; Lopes da Silva et al., 1974; Moran et al., 2007; Wendling et al., 2000; Zetterberg et al., 1978) and has been used recently as the basis of a generative model for event-related potentials that can be inverted using real data (David et al., 2006a; Friston et al., 2003; Jansen et al., 2001; Kiebel et al., 2006; Moran et al., 2007; Moran et al., 2008; Valdes et al., 1999).

In short, neural-mass models are special cases of ensemble density models that are furnished by ignoring all but the expectation or mode of the ensemble density. This affords the considerable simplification of the dynamics and allows one to focus on the behaviour of a large number of ensembles, without having to worry about an explosion in the number of dimensions or differential equations one has to integrate. An important generalisation of neural-mass models, which allow for states that are functionals of position on the cortical sheet, are referred to as *neural-field models* (see *Appendix* F; (Breakspear et al., 2006; Jirsa and Haken, 1996; Robinson et al., 2003;

Wright et al., 2003). (Deco et al., 2008) provide a comprehensive overview of neural-mass and neural field models, to which the interested reader is referred.

5.2.4 Summary

In conclusion, statistical descriptions of neuronal ensembles can be formulated in terms of a Fokker-Planck equation; an equation prescribing the evolution of a probability density on some phase-space. The high dimensionality and complexity of these Fokker-Planck formalisms can be finessed with a mean-field approximation to give nonlinear Fokker-Planck equations, describing the evolution of separable ensembles that are coupled by mean-field effects. By parameterising the densities in terms of their sufficient statistics, these partial differential equations can be reduced to ordinary differential equations describing the evolution of the statistics. In the simplest case, we can use a single statistic corresponding to the expectation or mode of the probability for each ensemble. This can be regarded as encoding the location of a probability mass. In what follows, we consider what would happen if the sufficient statistics included both the mean and dispersion.

5.3 Ensemble dynamics under the Laplace assumption

In this section, we derive a general mean-field reduction for neural dynamics formulated with any set of ordinary differential equations. This is formally equivalent to the method of moments (MM) proposed by (Rodriguez and Tuckwell, 1998; Rodriguez, 1996) for summarising density dynamics. In the next section, we apply the treatment to the equations used in dynamic causal modelling (DCM) of electrophysiological responses. The treatment here rests on summarising the ensemble density with a fixed form; namely, a Gaussian density. This corresponds to the Laplace assumption made in mean-field treatments in variational or ensemble learning in statistics. Here we use this approach to reduce a very high-dimensional integration problem into the manageable integration of the sufficient statistics (e.g., moments) of the ensemble density. The sufficient statistics are those quantities needed to define a

particular density, in this case the mean $\mu^{(i)}$ and covariance $\Sigma^{(i)}$ of the states of the *i*-th population, with a multivariate normal distribution; $q(x^{(i)}) = N(\mu^{(i)}, \Sigma^{(i)})$.

5.3.1 A single population

For simplicity, we will start with one population and generalise later. Consider some equations of motions for the dynamics of a single neuron and the corresponding density dynamics

$$\begin{aligned} \dot{x} &= f(x,u) + \Gamma(x) \\ \dot{q} &= -\nabla \cdot fq + \nabla \cdot D\nabla q \end{aligned} . \tag{5.7}$$

Here, we have introduced an exogenous input u that exerts its effect through the flow (*e.g.*, pre-synaptic input from another population causing a depolarisation and change in voltage). From these equations we can derive the equations of motion for the sufficient statistics (mean and covariance) of the ensemble density, $q(x) = N(\mu, \Sigma)$.

$$\begin{aligned} \dot{\mu}_i &= \int x_i \dot{q}(x) dx \\ &= \int f_i(x) q(x) dx \\ \dot{\Sigma}_{ij} &= \int \bar{x}_i \bar{x}_j \dot{q}(x) dx \\ &= \int (\bar{x}_j f_i(x) + \bar{x}_i f_j(x)) q(x) dx + D_{ij} + D_{ji} \end{aligned} . \tag{5.8}$$

Where $\bar{x}_i = (x_i - \mu_i)$ represent perturbations from the mean of the *i*-th state. These equalities can be verified using integration by parts; for example, with a single state we have

$$\begin{aligned} \dot{\mu} &= \int -x\partial_x (fq) dx + \int x\partial_x D\partial_x q dx \\ &= -\left[xf(x)q(x)\right]_{-\infty}^{\infty} + \int f(x)q(x)dx + \left[xD\partial_x q\right]_{-\infty}^{\infty} - \int D\partial_x q dx \\ &= \int f(x)q(x)dx \end{aligned} . \tag{5.9}$$

Here, we have used the fact that $q(x) = \partial_x q(x) = 0 : x \to \pm\infty$ is a proper density. The dynamics of the sufficient statistics in Eq. 5.8 are intuitively sensible; the rate of change of the mean of any state is the expected flow, in the direction of that state. Similarly, the variance only stops changing when dispersion due to random fluctuations is balanced by contraction due to flow. This contraction is proportional to the negative correlation between flow and the distance from the mean. This perspective can be made explicit by writing Eq. 5.8 as

$$\begin{aligned}\dot{\mu}_i &= \left\langle f_i(x) \right\rangle_q \\ \dot{\Sigma}_{ij} &= \left\langle \bar{x}_j f_i(x) + \bar{x}_i f_j(x) \right\rangle_q + D_{ij} + D_{ji}\end{aligned} \quad . \tag{5.10}$$

We can now exploit the fixed-form (Laplace) assumption about the ensemble density by rewriting Eq. 5.10 in terms of its sufficient statistics, using an expansion of the flow around the expected state

$$f_i(x) = f_i(\mu, u) + \sum_j \frac{\partial f_i}{\partial x_j} \bar{x}_j + \tfrac{1}{2} \sum_{jk} \frac{\partial^2 f_i}{\partial x_j \partial x_k} \bar{x}_j \bar{x}_k + \ldots . \tag{5.11}$$

Under Gaussian assumptions $\left\langle \bar{x}_i \right\rangle_q = 0$ and $\left\langle \bar{x}_i \bar{x}_j \right\rangle_q = \Sigma_{ij}$ and we get

$$\begin{aligned}\dot{\mu}_i &= f_i(\mu, u) + \tfrac{1}{2} \sum_{jk} \frac{\partial^2 f_i}{\partial x_j \partial x_k} \Sigma_{kj} \\ \dot{\Sigma}_{ij} &= \sum_k \frac{\partial f_i}{\partial x_k} \Sigma_{jk} + \sum_k \frac{\partial f_j}{\partial x_k} \Sigma_{ik} + D_{ij} + D_{ji}\end{aligned} \quad . \tag{5.12a}$$

This can be expressed more compactly in matrix form

$$\begin{aligned}\dot{\mu}_i &= f_i(\mu, u) + \tfrac{1}{2} tr(\Sigma \partial_{xx} f_i) \\ \dot{\Sigma} &= \partial_x f \Sigma + \Sigma \partial_x f^T + D + D^T\end{aligned} \quad . \tag{5.12b}$$

This is a key expression because it allows us to formulate population dynamics, under the Laplace assumption, knowing only the flow, its gradient and curvature ($f_i, \partial_x f_i, \partial_{xx} f_i$) at the expected state. Furthermore, we have circumnavigated the problem of integrating the density at every point in state-space to integrating a small number of sufficient statistics for each population. Equation 5.12 is instructive because it shows explicitly how the first and second moments of the density depend on each other; the variance affects the mean when and only when the curvature (second derivative) of the flow is non zero. This will always be the case if the equations of motion are nonlinear in the states. Similarly, the effect of the mean on the variance depends on nonlinear dynamics because the gradients in the second equality above will only change with the mean, when the curvature is non zero.

Interestingly, the form of neuronal dynamics implicit in Eq. 5.5 is linear in the states; in other words, $\partial_{xx} f_i = 0$. Equation 5.12 shows that the dynamics of the mean do not depend on the covariance and a neural-mass model is sufficient to model density dynamics. Below, we will consider a nonlinear conductance-based model where $\partial_{xx} f_i \neq 0$, which means there is a potential role for dispersion. Finally, Equation 5.12 shows that if we approximate the ensemble density with a point mass we recover the original equations of motion for a single neuron; *i.e.*, if $\Sigma = 0$ then the dynamics are completely specified in Eq. 5.12b by $\dot{\mu}_i = f_i(\mu, u)$. This is a *neural-mass model* and precludes interactions among moments of the population density.

5.3.2 Coupling different populations

Above, we treated each member of the neuronal population as evolving independently of the others, as if we were modelling a 'gas' of neurons. However, real neurons are connected and influence each other. We now consider mean-field equations for a set of m coupled populations that accommodate these influences. Under mean-field coupling each neuron 'senses' the states of all neurons in one or more populations. The ensuing effects can be formulated by making the motion of each neuron a function of population densities and, implicitly, their sufficient statistics, $\mu = \mu^{(1)}, \ldots, \mu^{(m)}$ and $\Sigma = \Sigma^{(1)}, \ldots, \Sigma^{(m)}$

$$\dot{x}^{(i)} = f(x^{(i)}, u, \mu, \Sigma) + \Gamma(x). \tag{5.13}$$

This couples the microscopic evolution of each neuron to macroscopic density dynamics within and between populations. These mean-field effects basically change the pattern of flow within a population's state-space. The corresponding density dynamics of the *j*-th population are now

$$\begin{aligned} \dot{\mu}_i^{(j)} &= f_i^{(j)}(\mu, \Sigma, u) + \tfrac{1}{2} tr(\Sigma^{(j)} \partial_{xx} f_i^{(j)}) \\ \dot{\Sigma}^{(j)} &= \partial_x f^{(j)} \Sigma + \Sigma \partial_x f^{(j)T} + D^{(j)} + D^{(j)T} \end{aligned}. \tag{5.14}$$

Notice that the terms involving gradients and curvatures pertain only to the population in question. This is because $\partial f^{(j)} / \partial x^{(i)} = 0 : \forall i \neq j$; in other words the motion in one population $f^{(i)} = f(x^{(i)}, \mu, \Sigma)$ depends only on the density on the states of others, not the states *per se*. Before turning to a specific example we consider the outputs or responses of these systems.

5.3.3 Observed responses

In the next Chapter, we will use the density dynamics above as the basis of a dynamic causal model (DCM) of observed data. This requires one to specify how the density maps to observed responses, such as the electroencephalogram (EEG) or blood oxygen level dependent (BOLD) signals in functional magnetic ressonance imaging (fMRI). Generally, these observations are generated by an average $\eta(\mu, \Sigma) = \left\langle g(x^{(i)}) \right\rangle_q$ of some nonlinear function of the states, $g(x^{(i)})$. The average is usually over millions of neurons in an assumed electromagnetic source or voxel in neuroimaging and is a function of and only of the sufficient statistics. For EEG this function may simply scale the depolarization of pyramidal cells (*e.g.*, $g(x^{(i)}) = g x_j^{(i)}$); we will use this below. For fMRI $g(x^{(i)}) = H(x_j^{(i)} - \tau)$ may be a Heaviside or threshold function of depolarisation to reflect synaptic firing. Under the Laplace

assumption the expected firing rate $\eta(\mu,\Sigma)$ becomes a sigmoid [error] function of the mean depolarisation.

5.3.4 Application to a conductance-based model

In this section, we apply the Laplace approximation to a model we have used in previous papers (Friston et al., 2003), which has a complexity intermediate between simple integrate-and-fire models and Hodgkin-Huxley models. Conductance-based models are the most common formulation used in neuronal models and can incorporate as many different ion channel types as are known for the particular cell being modelled. Some examples of conductance-based models are Hodgkin-Huxley model (1952), Connor-Stevens model (1971), Morris-Lecar model (1981). This involves specifying the equation of motion and implicitly their gradients and curvatures. These quantities specify the density dynamics in terms of sufficient statistics under the Laplace assumption. Finally, we will look at some special cases that will be compared in the final section of this Chapter.

5.3.4.1 *The equations of motion*

The neuronal dynamics of any given population considered here conform to a simplified (Morris and Lecar, 1981) model, where the states $x^{(i)} = \left\{V^{(i)}, g_1^{(i)}, g_2^{(i)}, \ldots\right\}$ comprise transmembrane potential and a series of conductances corresponding to different types of ion channel. The dynamics are given by the stochastic differential equations

$$\begin{aligned} C\dot{V}^{(i)} &= \sum_k g_k^{(i)}(V_k - V^{(i)}) + I + \Gamma_V \\ \dot{g}_k^{(i)} &= \kappa_k^{(i)}(\varsigma_k^{(i)} - g_k^{(i)}) + \Gamma_k \end{aligned} \quad . \tag{5.15}$$

These equations of motion constitute a model for a single neuron and, when solved simultaneously for an ensemble of neurons, furnish an ensemble model. They are effectively the governing equations for a parallel resistance-capacitance circuit; the

first says that the rate of change of transmembrane potential (times capacitance, C) is equal to the sum of all currents across the membrane (plus exogenous current, $I = u$). These currents are, by Ohm's law, the product of potential difference between the voltage and reversal potential, V_k for each type of conductance. These currents will either hyperpolarise or depolarise the cell, depending on whether they are mediated by inhibitory or excitatory receptors respectively (*i.e.*, whether V_k is negative or positive). Conductances change dynamically with a characteristic rate constant κ_k and can be regarded as the number of open channels. Channels open in proportion to pre-synaptic input ς_k and close in proportion to the number open. The pre-synaptic input corresponds to the expected firing rate in another population, times a coupling parameter γ_{ij}^k for the *k*-th conductance

$$\begin{aligned}\varsigma_k^{(i)} &= \sum_j \gamma_{ij}^k \int q(V^{(j)}) H(V^{(j)} - V_R) dV^{(j)} \\ &= \sum_j \gamma_{ij}^k \sigma(\mu_V^{(j)} - V_R, \Sigma^{(j)})\end{aligned} \quad . \tag{5.16}$$

where $H(\cdot)$ is a Heaviside function and the sigmoid function $\sigma(\cdot)$ is a cumulative density on the depolarisation; see Chapter 3 and Equation 5.17 below. The form of Equation 5.16 is motivated in detail in Chapter 4.

The coupling parameters specify connectivity among populations. Furthermore, they can be used to ensure that each population couples to one and only one conductance type (*i.e.*, each population can only release one sort of neurotransmitter). Generally, one would model a neuronal network of areas, where each area comprises two or more populations. This engenders the distinction between intrinsic and extrinsic connections, which couple populations within and between brain areas. In this Chapter, we restrict ourselves to a single area and intrinsic connections; however, there is no mathematical distinction between intrinsic and extrinsic connections. The firing in source populations is a Heaviside or threshold function of depolarization where the threshold, V_R determines the proportion of afferent cells firing. Under the mean-field assumption, this input is a function of the population density of the source

and, under the Laplace assumption, this function is simply the Gaussian cumulative density

$$\sigma(\mu,\Sigma) = \left(2\pi \det(\Sigma)\right)^{-\frac{1}{2}} \int_{-\infty}^{\mu} \exp\left(-\tfrac{1}{2} x^T \Sigma^{-1} x\right) dx. \tag{5.17}$$

and is a function of the source's sufficient statistics. These equations constitute $f^{(i)} = f(x^{(i)}, u, \mu, \Sigma)$ of the previous section and are sufficient to elaborate a mean-field approximation under the Laplace assumption using Eq. 5.14; where (dropping the population superscript for clarity)

$$f = \begin{bmatrix} \frac{1}{C}\sum_k g_k (V_k - V) + \frac{1}{C} I \\ \kappa_1(\varsigma_1 - g_1) \\ \kappa_2(\varsigma_2 - g_2) \\ \vdots \end{bmatrix}$$

$$\partial_x f = \begin{bmatrix} -\frac{1}{C}\sum_k g_k & \frac{1}{C}(V_1 - V) & \frac{1}{C}(V_2 - V) & \cdots \\ 0 & -\kappa_1 & 0 & \\ 0 & 0 & -\kappa_2 & \\ \vdots & & & \ddots \end{bmatrix} \tag{5.18}$$

$$\partial_{xx} f_V = \begin{bmatrix} 0 & -\frac{1}{C} & -\frac{1}{C} & \cdots \\ -\frac{1}{C} & 0 & 0 & \\ -\frac{1}{C} & 0 & 0 & \\ \vdots & & & \ddots \end{bmatrix}$$

$$\partial_{xx} f_g = 0$$

Note that the curvature has a simple form because the equations of motion are second order only in voltage and conductance. An example of the expressions for the ensuing motion of the sufficient statistics $\lambda^{(i)} = \{\mu^{(i)}, \Sigma^{(i)}\}$ from Equation 5.14 and the corresponding Jacobian, $\partial_\lambda \dot{\lambda}$ are provided in Figure 5.1, for two populations. This figure provides an iconic summary of how different quantities affect each other. For

example, the variances affect only the mean depolarisation, in inverse proportion to the capacitance. These equations are not necessary to integrate the sufficient statistics; we only derived the subset of equations shown in the figure for didactic purposes. In practice, these derivatives are evaluated numerically, given the user specified equations of motion.

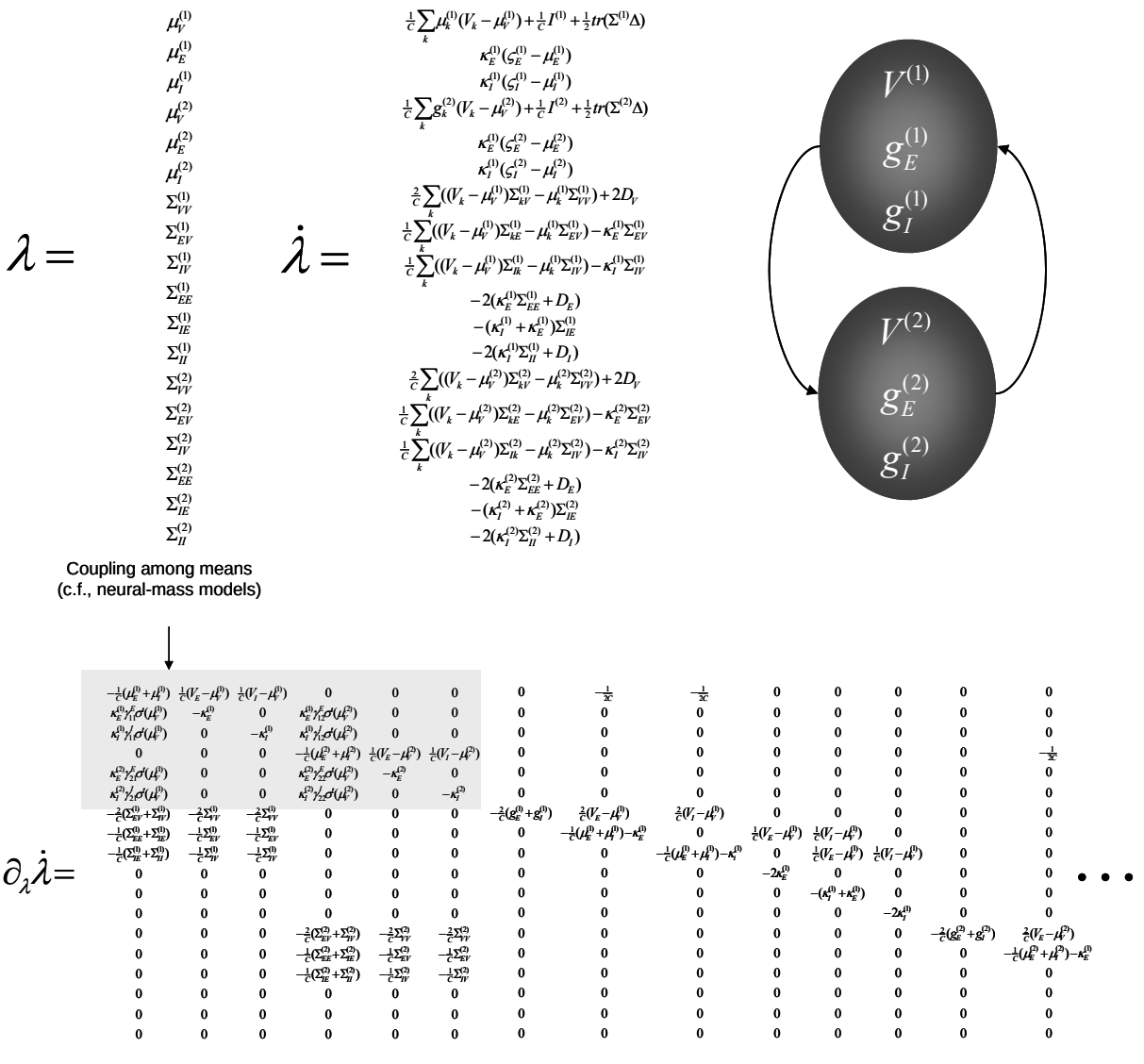


Figure 5.1: Expressions for the motion of the sufficient statistics $\lambda^{(i)} = \mu^{(i)}, \Sigma^{(i)}$ (mean and variance) and the corresponding Jacobian for two populations that conform to simplified Morris-Lecar-like dynamics. The grey area in the Jacobian covers terms that link mean states to each other and are considered in neural-mass reductions of full mean-field models. The equations are only used iconically.

5.3.4.2 *Some special cases*

Before assessing the accuracy of the Laplace scheme we will consider some special cases of Equation 5.14. The first is obtained if we assume $\Sigma^{(i)}$ is fixed for all populations. Because the covariance is fixed, we only have to integrate the ensemble mean; furthermore because the curvature is constant (voltage) or zero (conductance), this entails an extra decay term for voltage, giving density dynamics of the form

$$\begin{aligned}\dot{\mu}_V^{(i)} &= f_V^{(i)}(\mu,\Sigma,u) + \tfrac{1}{2}tr(\Sigma^{(i)}\partial_{xx}f_V^{(i)}) \\ \dot{\mu}_g^{(i)} &= f_g^{(i)}(\mu,\Sigma,u)\end{aligned}. \tag{5.19}$$

This corresponds to a neural-mass model with decay and will be used for comparative analysis in the next section. Finally if we further assume that $\Sigma^{(i)}$ is spherical (i.e., all off-diagonal terms are zero) then this decay terms disappears because the leading diagonal of $\partial_{xx}f_V^{(i)}$ is zero. In this instance, the dynamics reduce to the original equations of motion because we can ignore the second-order statistics completely

$$\dot{\mu}_k^{(i)} = f_k^{(i)}(\mu,\Sigma,u). \tag{5.20}$$

This is a conventional neural-mass model with the usual sigmoid activation function. This function depends on the variance (see Eq. 5.15), which we assume is fixed. Note that this provides another perspective on the parameterisation of the sigmoid activation function in classical neural-mass models (c.f. Eq. 5.5 and the derivations in Chapter 4). In the next section we will compare the Laplace (Eq. 5.14) and neural-mass approximations (Eq. 5.19) in terms of modelling evoked neuronal transients.

5.4 Summary

We are now in a position to compare and contrast ensemble models of neuronal populations with mean-field (MFM) and neural-mass (NMM) approximations. Ensemble models (Eq. 5.15) provide the trajectories of many neurons to form a sample density of population dynamics. The MFM is obtained by a mean-field and a

Laplace approximation to these densities (Eq. 5.14). The NMM is a special case of the mean-field model in which we ignore all but the first moment of the density (*i.e.*, the mean or mode). In other words, the NMM discounts dynamics of second-order statistics (*i.e.*, variance) of the neuronal states. The mean-field models allow us to model interactions between the mean of neuronal states (e.g., firing rates) and their dispersion or variance over each neuronal population modelled (c.f., (Harrison et al., 2005)). The key behaviour we are interested in is the coupling between the mean and variance of the ensemble, which is lost in the NMM. The different models and their mathematical representations are summarised in Table 5.1.

Model	*Description*	*Equation*
Ensemble	*Stochastic differential equation* that describes how the states evolve as functions of each other and some random fluctuations	$dx = f(x,u)dt + \sigma dw$ (Eq.5.2)
MFM	*Differential equation* that describes how the density evolves as functions of mean and covariance. Resulting from a *mean-field* and *Laplace* approximations of the ensemble model	$\dot{\mu}_i^{(j)} = f_i^{(j)}(\mu,\Sigma,u) + \frac{1}{2}tr(\Sigma^{(j)}\partial_{xx}f_i^{(j)})$ $\dot{\Sigma}^{(j)} = \partial_x f^{(j)}\Sigma + \Sigma\partial_x f^{(j)T} + D^{(j)} + D^{(j)T}$ (Eq.5.14)
NMM	*Differential equation* that describes how the density evolves as a function of the mean. Obtained by *fixing* the co*variance* of the MFM	$\dot{\mu}_i^{(j)} = f_i^{(j)}(\mu,\Sigma,u) + \frac{1}{2}tr(\Sigma^{(j)}\partial_{xx}f_i^{(j)})$ $\dot{\Sigma}^{(j)} = 0$ (Eq.5.19)

Table 5.1: Overview of the three models: Ensemble, Mean-Field model (MFM) and Neural-Mass model (NMM). For a detailed description of the equations see main text.

5.5 Neural-mass *vs.* mean-field models

In this section, we examine the accuracy of the Laplace approximation to density dynamics, in relation to the true density dynamics that obtain by integrating the trajectories of a real but finite-sized population. We will also take the opportunity to highlight the difference between the Laplace approximation and neural-mass simplifications. In what follows, we examine the response of three populations connected to emulate the source model for electromagnetic responses we use in DCM for ERPs (David et al., 2006a; Kiebel et al., 2006). Each electromagnetic source comprises two excitatory populations and an inhibitory population. These are taken to represent input cells (spiny stellate cells in the granular layer of cortex), inhibitory interneurons (allocated somewhat arbitrarily to the superficial layers) and output cells (pyramidal cells in the deep layers). The deployment and intrinsic connections among these populations are shown in Figure 5.2 and the parameters are provided in Table 5.2.

Parameter	*Physiological Interpretation*	*Value*
g_L	Leaky conductance	*1 mV*
$\tau_{E,I} = 1/\kappa_{E,I}$	Postsynaptic rate constants	*4ms, 16ms*
$\gamma_{31}^E, \gamma_{13}^E, \gamma_{23}^E, \gamma_{12}^I, \gamma_{32}^I$	Intrinsic connectivity	*1, 0.5, 1, 0.5, 2*
V_L, V_E, V_I	Reversal potential	*-70 mV, 60 mV, -90 mV*
V_R	Threshold potential	*-40 mV*

Table 5.2: Parameter values for all models used in this Chapter.

In this model, we use three conductance types: leaky, excitatory and inhibitory conductance. This gives, for each population

$$
\begin{aligned}
C\dot{V}^{(i)} &= g_L(V_L - V^{(i)}) + g_E^{(i)}(V_E - V^{(i)}) + g_I^{(i)}(V_I - V^{(i)}) + I + \Gamma_V \\
\dot{g}_E^{(i)} &= \kappa_E(\varsigma_E^{(i)} - g_E^{(i)}) + \Gamma_E \\
\dot{g}_I^{(i)} &= \kappa_I(\varsigma_I^{(i)} - g_I^{(i)}) + \Gamma_I \\
\varsigma_k^{(i)} &= \sum_j \gamma_{ij}^k \sigma(\mu_V^{(j)} - V_R, \Sigma^{(j)})
\end{aligned}
\quad . \qquad (5.21)
$$

Notice that the leaky conductance does not change, which means the states reduce to $x^{(i)} = \left\{V^{(i)}, g_E^{(i)}, g_I^{(i)}\right\}$. Furthermore, for simplicity, we have assumed that the rate-constants, like the reversal potentials are the same for each population. The excitatory and inhibitory nature of each population is defined entirely by the specification of the non-zero intrinsic connections γ_{ij}^k (see Figure 5.2). The resulting sparse connectivity means that not all populations have all conductances.

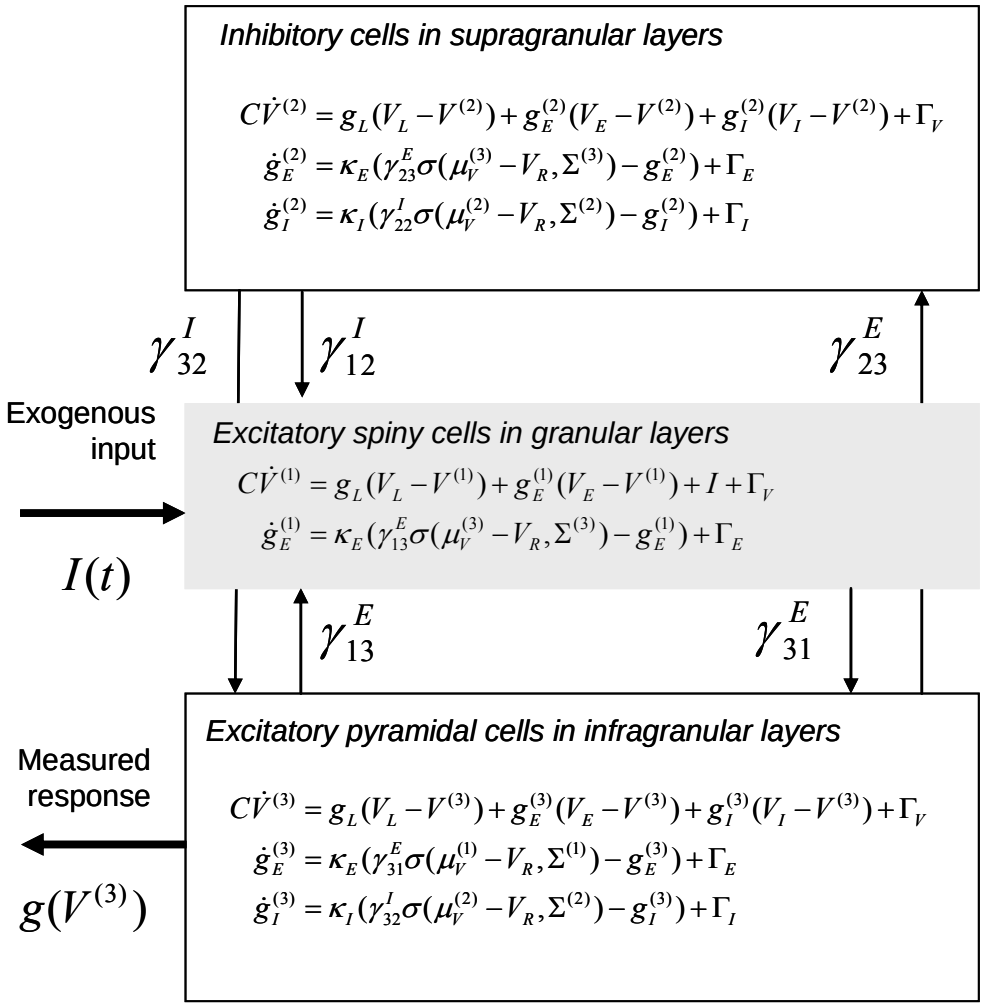


Figure 5.2: Neuronal state-equations for a source model with a layered architecture comprising three interconnected populations (Spiny-stellate, Interneurons, and Pyramidal cells), each of which has three different states (Voltage, Excitatory and Inhibitory conductances).

5.5.1 Simulations

In what follows, we examine the response of this three-population source to an exogenous input using the Laplace and neural-mass approximations. We first compare the analytic approximations based on the mean-field (Eq. 5.14), with the sample density of responses from simulated neuronal ensemble (Eq. 5.15). We present more comprehensive characterisations, comparing predicted responses under mean-field and neural-mass models to transient and sustained input. Our aim was to (i) evaluate the Laplace approximation in relation to the response obtained by integrating the original stochastic equation of motions and (ii) to compare the Laplace approximation (Eq.5.14) with the neural-mass model (Eq. 5.19) to assess the need for population covariance as part of the model.

5.5.2 Ensemble dynamics

In the first simulations, we examined population responses to an impulse or burst of afferent input. This can be regarded as a simple evoked response. We integrated the equations of motion (Eq. 5.15) for the three population model of Figure 5.2, with 64 neurons per population. To integrate the stochastic differential equations, we added a random normal variate to the states of each neuron, at each time step Δt sampled from a Gaussian density with variance, $2D\Delta t$. The ensuing impulse responses are shown in Figure 5.3, in terms of the depolarisation of pyramidal cells. Because we used a relatively small ensemble of neurons there are some (but not marked) finite-size effects: Finite-size effects are seen when approximating the response of a large ensemble with the response of a small number of neurons (see (Doiron et al., 2006; Galan et al., 2007; Mattia and Del Giudice, 2004) for a discussion of finite element methods in characterising the behaviour of neuronal ensembles). Critically, the random fluctuations due to the Wiener processes lead to different trajectories (Figure

5.3; middle panel), which provide a sample density for the population dynamics. This can be summarised in terms of its mean and a 90% confidence interval, over peristimulus time (Figure 5.3; lower panel). The key thing to observe here is that the dispersion is not stationary; it changes with time. Specifically, when the states are changing quickly around the peak response, the dispersion of states is much smaller than when the ensemble is at baseline. It is this change in dispersion that is discounted by conventional NMMs.

Figure 5.4 shows the integrated responses of this ensemble of neurons, for all states and populations. The red arrows show the main causal influences that couple different populations. These are the mean-field effects of depolarisation in one population increasing the excitatory or inhibitory conductance of another (through intrinsic connections). This, in turn leads to depolarisation or hyperpolarisation of the target population. The configuration of intrinsic connections means that input, which enters at the spiny stellate population, may only be expressed ten or more milliseconds later in other populations. It is these slow population effects we want to approximate.

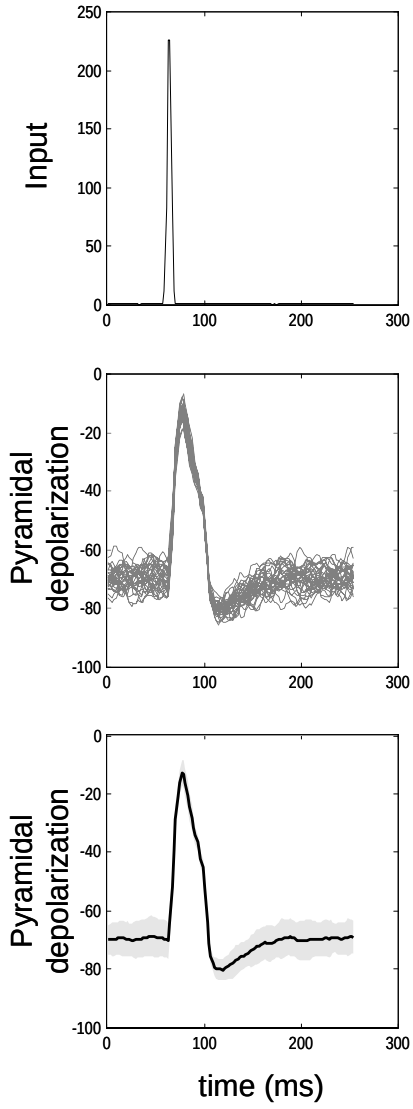


Figure 5.3 Top: Exogenous input. Middle: Integrated response (64 neurons) of the pyramidal population, where the spike is driving the neuronal source through intrinsic connections (Figure 5.2). Bottom: Summary of the density over trajectories in terms of their mean (solid line) and a 90% confidence interval (grey region).

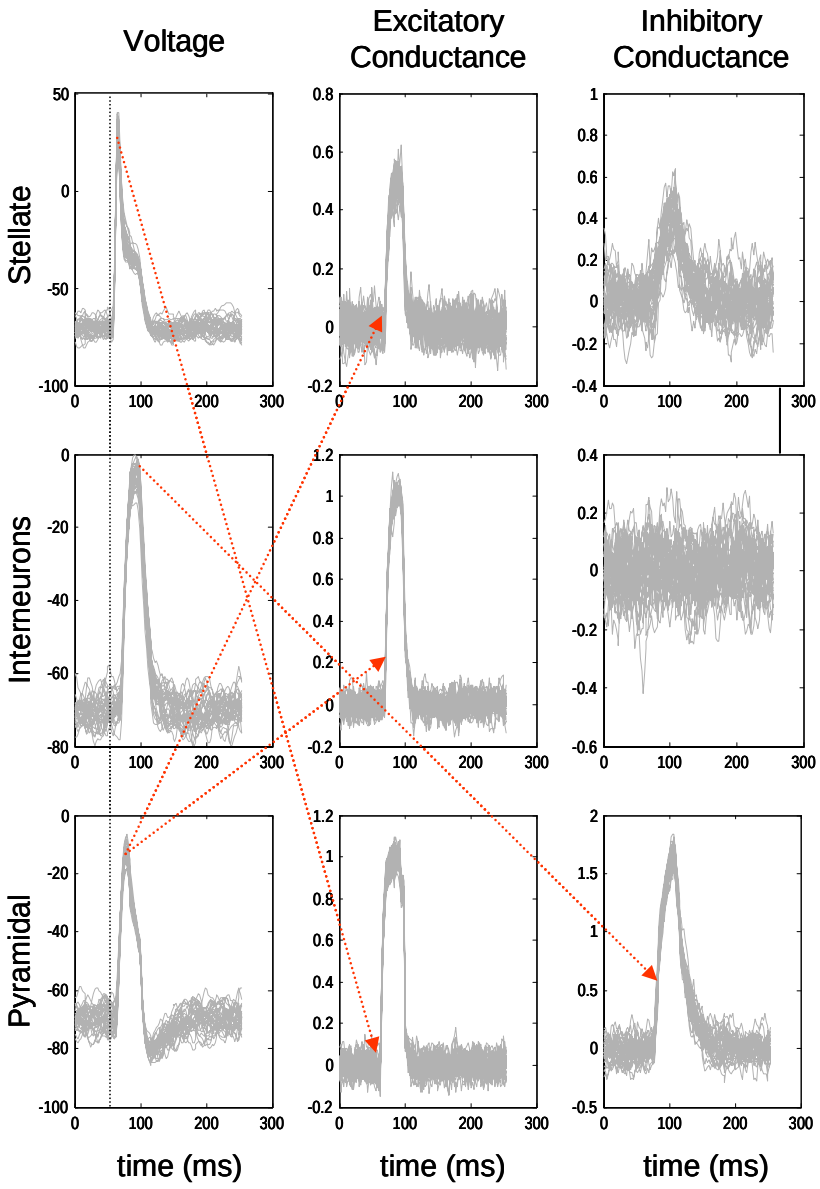


Figure 5.4 Ensemble model responses for the three neuronal populations (stellate, interneurons, pyramidal) over their three different states (voltage, excitatory and inhibitory conductance). The red lines correspond to the causal influences mediated by intrinsic connections that convey mean-field effects (from voltage to conductances). The vertical broken line is aligned to the exogenous input that arrives at 64 ms.

We solved the population dynamics to give the MFM and the NMM approximations to the impulse responses in Figure 5.4. For the MFM, the mean and dispersion of the state dynamics were computed by solving Eq. 5.14 for the same model and input used above. The NMM dynamics were obtained by fixing the dispersion of the MFM to its steady-state value (in the absence of input); this is the stationary solution to Eq. 5.14. There have been no previous attempts to quantify the difference between the derived NMM and MFMs beyond one appearing to resemble the original ensemble dynamics more closely than the other. In Fig. 5.5 we compare the output of the three described models for two different neuronal source inputs. One can see that the mean of the trajectories are similar for all models. Although one can see that after the peak, the mean response of the MFM response is more like the ensemble model than the NMM response. Furthermore, like the ensemble model, the MFM dispersion changes over time, while the dispersion of the NMM is constant. For more complex source models these small differences may have significant repercussions, if the dynamics of the mean depend on dispersion. We will see an example of this later. The MFM appears to overestimate the dispersion in comparison to the ensemble model; however, this is probably due to finite size effects.

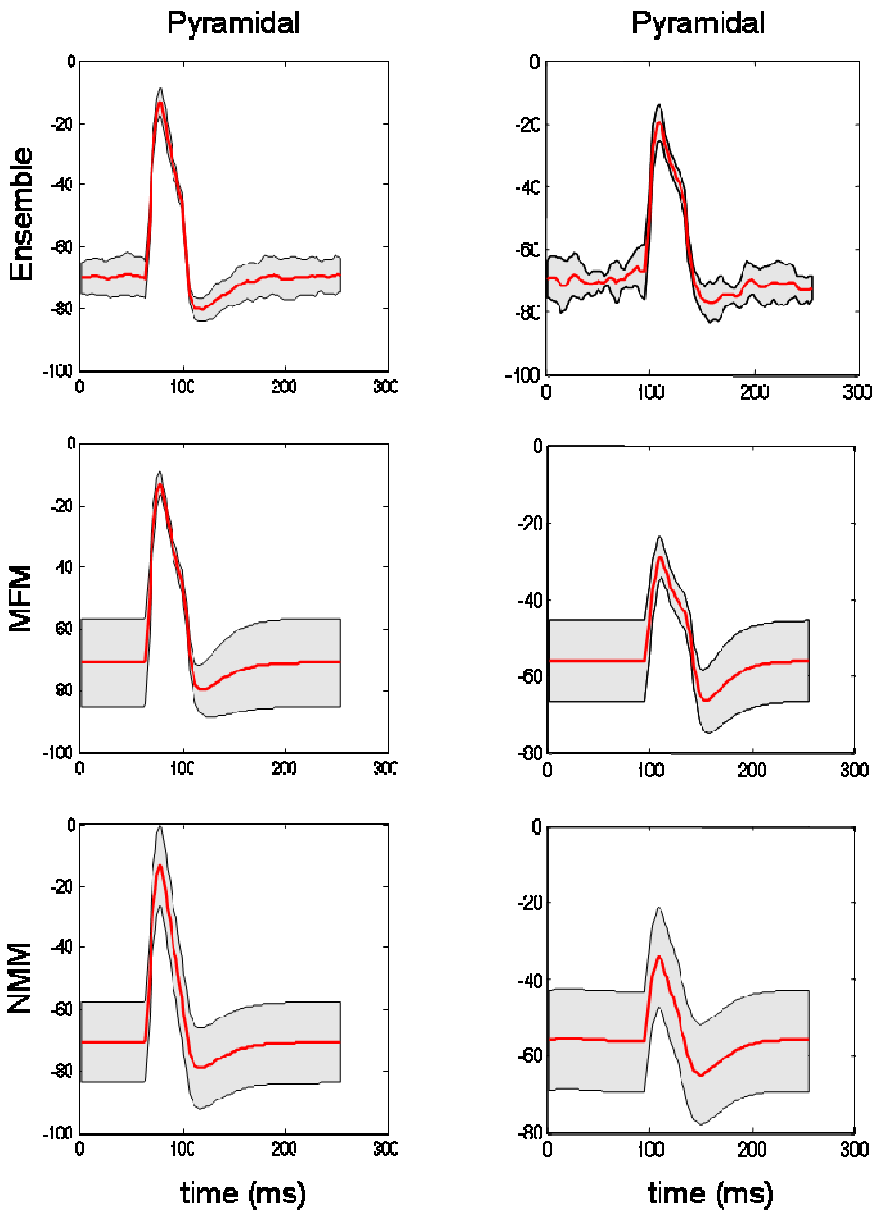


Figure 5.5: Population response of the pyramidal cells for the three models: ensemble model, mean-field model and neural-mass model. One can see differences for the mean (solid lines) and the dispersion (grey regions) of the trajectories. See Figure 5.2 for the neuronal state-equation source model.

5.5.3 Comparing MFM and NMM predictions

Using the above model, we compared the MFM and the NMM responses using exogenous inputs that varied in amplitude and were transient or sustained. The results of these simulations are shown in Figure 5.6, in terms of pyramidal cell population depolarisation. With transient inputs we found that both the MFM and the NMM predicted a similar response. The two models respond with a short-lived burst of activity that increased with input amplitude and showed a plateau around 60 μA. When the input exceeds 90 μA, the response under both models become biphasic, with a second peak that lasted for about 30 ms. With sustained input, both models show complex nonlinear behaviour for input amplitudes greater than 24 μA. However, for input amplitude values greater than 50 μA the response patterns of the two models are very different. The MFM shows a sustained oscillatory or limit-cycle behaviour that is largely unaffected by further increases in input. In contrast, the NMM returns to a fixed level of depolarization (a fixed-point attractor) after about 200 ms; this illustrates that the MFM retains key nonlinearities and can exhibit bifurcations that are structurally distinct from the NMM. In short, one observes subtle but potentially important differences between the two models, which may have important implications for generative models of observed neuronal responses.

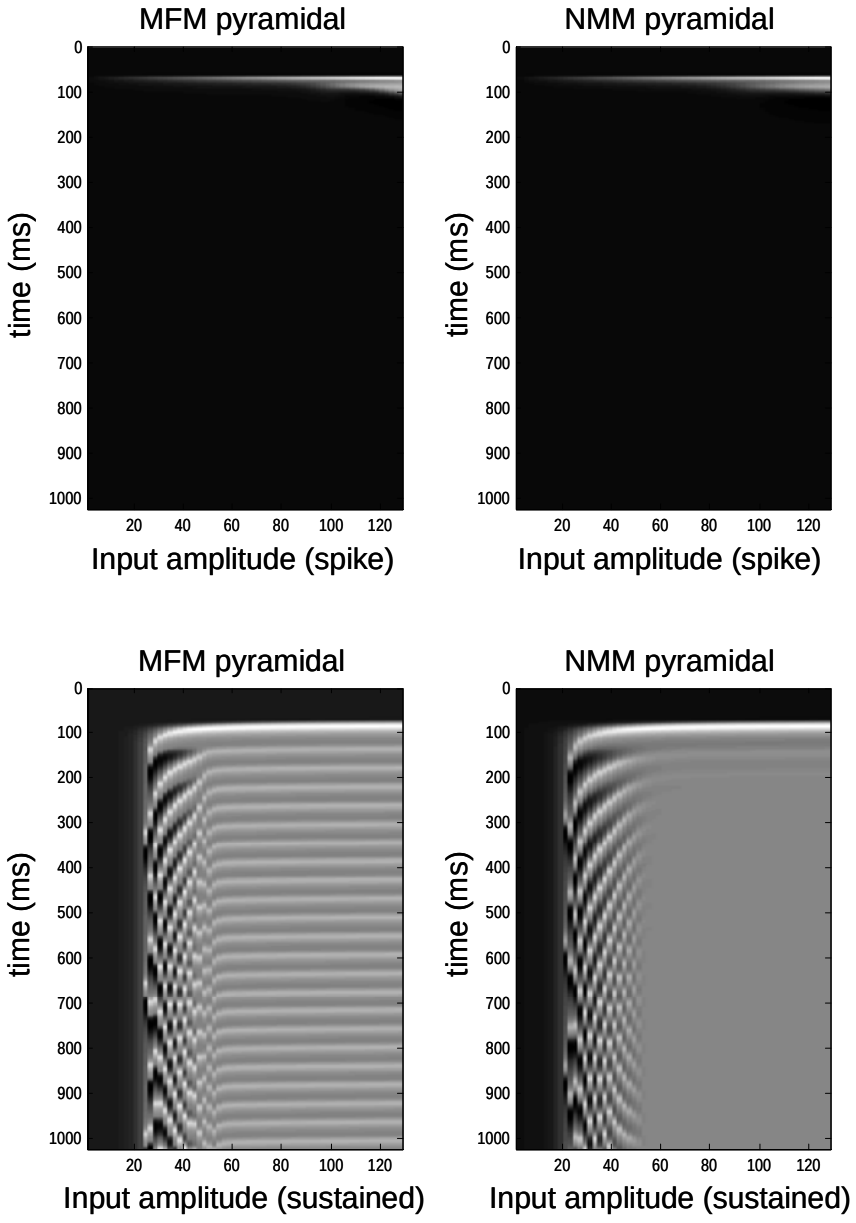


Figure 5.6: (**Left**): Pyramidal population response (depolarization) under the mean-field model to varying levels of input. (**Right**): Equivalent pyramidal population response under the neural-mass model. (**Top row**) transient input at 64 ms; (**Lower row**) sustained input. The key thing to note is the difference between the predictions of the two models in the lower panels, which show the mean-field model prediction to oscillate at high levels of input. White indicates -10 mV and black -80 mV.

5.5.4 A quantitative characterisation

To quantify neuronal responses to sustained input under the MFM, we used frequency analyses and mean spiking responses. We focussed on the pyramidal population, which represents the principal (output) cells in cortex and are the predominant source of electromagnetic signals that are observed empirically. The results of these analyses are shown in Figure 5.7 using the same model and range of sustained input as above. It can be seen that the spectral responses are greatest between about 8 and 16 Hz, for input amplitudes between 25 and 45 μA (Figure 5.7A). In this range, the peak frequency increases almost linearly with amplitude. In Figure 5.7B we look in more detail at the MFM spectral response profile at input amplitudes of 32 and 64 μA. These two input levels fall into two different regimes of the spectral response (broken lines in Figure 5.7A). For the 32 μA input there is a pronounced alpha peak at ~10Hz, for the 64 μA input, the spectrum has a small beta peak around 24 Hz.

We next looked at how the population firing response probability scales with input amplitude. Figure 5.7C shows that a response emerges, after about 100 ms, at about 25 μA input amplitude and shows nonlinear behaviour over time; for higher input amplitudes, the activity oscillates at a constant frequency. This response pattern is very similar to the depolarization (Figure 5.6), because firing rate is a nonlinear function of the density on pyramidal depolarization. The ensuing input-firing rate curve (averaged over peristimulus time) shows a highly nonlinear behaviour, with no firing below a threshold of 20 μA and progressive increases until the firing saturates at input amplitudes of about 50 μA (Figure 5.7D).

This sort of simulation demonstrates that the limit-cycle attractor of the MFM can be exploited to study the relationship between oscillatory dynamics and mean levels of firing. In this instance, the model suggests that high firing rates, induced by sustained inputs, will be expressed in the context of higher frequencies in a desynchronised or ‘activated’ EEG. This is entirely consistent with empirical observations (e.g., (Kilner et al., 2005) and references therein). More generally, this simple simulation shows that the nature of responses predicted by mean-field and neural-mass models of exactly the same neuronal system can differ profoundly in terms of the dynamics they support. Here, the addition of extra variables encoding population covariance leads to

oscillations, under sustained input that are not predicted by a reduced neural-mass model. In principle, this means that mean-field DCMs of evoked and induced responses may provide better models of empirical data. We pursue the theme of nonlinearity and limit-cycles in the final simulations, which look at nested oscillations.

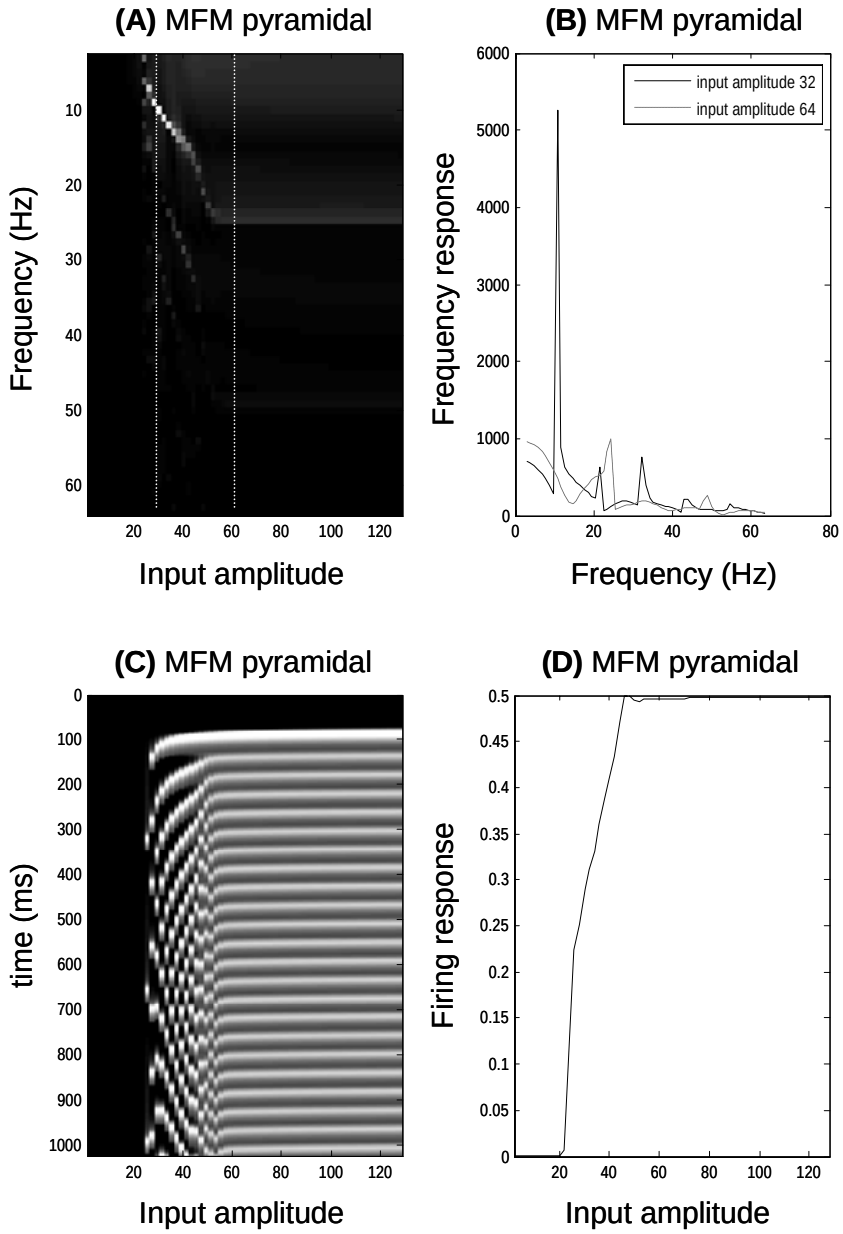


Figure 5.7: Mean-field model frequency response for the pyramidal population. (**A**) Spectral density of response as a function of input amplitude; (**B**) Spectral density of

response for input amplitude of 32 and 64 μA (broken lines in **A**); (**C**) Pyramidal firing rates as a function of time and input amplitude; (**D**) Mean population firing over time as a function of input amplitude.

5.5.5 Modelling nested oscillations and phase-synchronisation

Nonlinear coupling between distinct brain regions are observed, predominantly as interactions between low and high frequencies. These nonlinear influences are thought to mediate top-down modulation, 'attentional' and other context-defining functions (Canolty et al., 2006; Kopell et al., 2000; Varela et al., 2001; von Stein et al., 2000). Two principal forms of cross-frequency phase interactions have been recognized: '*n:m* phase synchrony', which indicates amplitude-independent phase-locking of *n* cycles of one oscillation to *m* cycles of another oscillation (Palva et al., 2005; Tass et al., 1998); and 'nested oscillations', which reflect the locking of the amplitude fluctuations of faster oscillations to the phase of a slower oscillation (Canolty et al., 2006; Penny et al., 2008; Vanhatalo et al., 2004). Nested oscillations have been observed in both the human brain and rat hippocampus (Chrobak and Buzsaki, 1998; Mormann et al., 2005); they have been proposed to underlie the discrete nature of perception and the capacity of working memory (Penny et al., 2008), as well as playing a role in sleep (Steriade, 2006) and olfaction (Kepecs et al., 2006). There many studies which rest on cross-frequency coupling, for example (Fukai, 1999; Haenschel et al., 2007; Hocking, 2007; Lisman and Idiart, 1995).

Motivated by these findings, we reproduced nested oscillations using our three-population source (Figure 5.2). We drove the neuronal source with a slow sinusoidal input to elicit periods of bursting in the inhibitory population. This produced phase-amplitude coupling, most notably between the inhibitory population and the spiny population that was driven by the low-frequency input. The bursting and concomitant nested oscillations are caused by nonlinear interactions between voltage and conductance, which are augmented by coupling between their respective means and dispersions. Figure 5.8 shows the predicted responses from the MFM and NMM models. The population responses of the MFM and NMM show clear differences in the number and amplitude of the oscillations per cycle of the low frequency input.

Again this illustrates the potential importance of using a MFM (as opposed to a NMM) when modelling nonlinear or quasi-periodic dynamics, like nested oscillations. This simulation is another illustration of how small differences between models can have large effects on the nature of predicted neuronal responses.

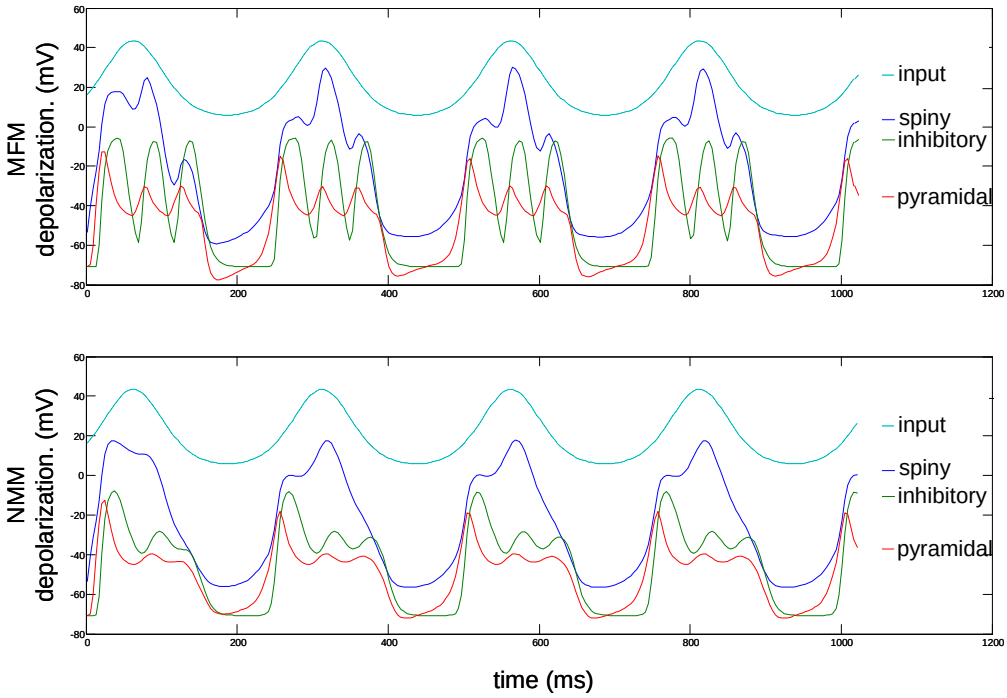


Figure 5.8 Nested oscillations in the three-population source driven by slow sinusoidal input for both MFM and NMM. Input is shown in light blue, spiny interneuron depolarization in dark blue, inhibitory interneurons in green and pyramidal depolarization in red. The nonlinear interactions between voltage and conductance produces phase-amplitude coupling in the ensuing dynamics. The MFM shows deeper oscillatory responses during the nested oscillations.

5.6 Discussion

The purpose of this work was to describe a generic approach to modelling dynamics in neuronal populations. Our work is motivated by the observation that neural-mass approaches, currently used as generative models for observed data (David et al., 2006a), are a limiting case of mean-field models. In other words, they consider only the first moment of the density for each population, which is a special case of the more general ensemble density formulation. In this Chapter, we augmented the neural-mass model with quantities that encode population dispersion to furnish mean-field models that capture full density dynamics.

The high dimensionality and complexity of Fokker-Planck formalisms can be reduced with a mean-field approximation, which describes the evolution of separate ensembles coupled by mean-field effects. By parameterising the densities in terms of their sufficient statistics, the partial differential equations can be reduced to ordinary differential equations describing the evolution of its sufficient statistics or moments (Table 5.2). In this way, we obtained a key equation (Eq. 5.14), which formulates population dynamics, using only the flow, its gradient and curvature, at the mean state. This expression shows explicitly how the first and second moments of the density depend on each other; the variance affects the mean if and only if the curvature (second derivative) of the flow is non zero. This will be the case if the equations of motion are nonlinear in the neuronal states. Similarly, the effect of the mean on the variance depends on nonlinear dynamics because the gradients will only change with the mean, when the curvature is non zero.

We have looked at the neuronal response of a particular but ubiquitous model (Figure 5.2) in terms of the mean and the dispersion of its underlying neuronal states (Figures 5.3 and 5.4). We established the validity of the Laplace approximation by comparing the response of a simulated ensemble of neurons to the response under the Laplace and neural-mass assumptions. The key behaviour we were interested in was the coupling between the mean and variance of the ensemble, which is lost in the neural-mass approximations. This enabled us to compare equivalent mean-field and neural-mass models of the same populations and evaluate, quantitatively, the contribution of

population variance to shaping population dynamics. The simulations for the Laplace mean-field model, which considers second-order statistics, support a more realistic and plausible model than the neural-mass model. The MFM shows, for an impulse response function, a dynamical behaviour that is more similar than the NMM to the response obtained by integrating the stochastic ensemble dynamics (Figure 5.5). Although the NMM is used widely because of its simplicity, it only considers the mean neuronal state and does not consider higher statistics like the variance. We speculate that this simplifying assumption may have implications when trying to invert generative models of real data.

The particular form of neuronal model used here (see Eq. 5.21 and Figure 5.2) is among the simplest that are nonlinear in the states (note that the rate of change of voltage depends on conductance times voltage). This nonlinearity is critical in the present context because, as discussed above, in its absence there is no coupling between the mean and dispersion (i.e., the neural mass and mean field formulations would behave identically). We are not suggesting, in the choice of this model, that it is a sufficient or complete model of neuronal dynamics; this would be a wider question for model comparison. We are using this minimal model to ask whether mean field formulations provide, in principle, a better account of observed neuronal responses than their neural mass counterparts. Moreover, note that the complexity of the NMM and MFM are the same; the MFM has more states but does not have more unknown parameters (see Table 5.1). This may seem counterintuitive because the dispersion in the MFM may appear to make it more complicated. However, the dispersion is a sufficient statistic of a density on hidden states and is not itself subject to random effects. This means, given the model parameters, it is a deterministic quantity and does not add to model complexity (i.e., it is specified by the same parameters as the neural mass model).

We compared the Laplace approximation (Eq. 5.14) with the neural-mass model (Eq. 5.19) to assess the role of the population covariance. NMMs, despite their relative simplicity, exhibited complex dynamical behaviour reminiscent of real neuronal responses. However, qualitative differences between MFM and NMM predictions were easy to demonstrate. In particular, we saw that the MFM showed a bifurcation

from fixed-point to a limit-cycle attractor, as sustained input levels were increased (Figure 5.6). We also looked at the spectral response of the pyramidal population of the mean-field model (Figure 5.7). This analysis disclosed the presence of physiologically plausible oscillatory signals in the alpha and beta band and how their relative power changed with activation. Additionally, we presented an interesting example of the quantitative difference between MFM and NMM by reproducing nested oscillation behaviour (Figure 5.8). In short, the MFM appeared to represent richer and more complex dynamics. This approach may have potential applications in dynamic causal modelling of imaging studies (M/EEG, fMRI) where one tries to explain the coordinated activity of a large number of neurons.

The Laplace assumption is a common device in statistical physics and finesses the problem of integrating a complicated density by assuming a Gaussian form. In machine learning, it allows one to focus on its sufficient statistics, namely the mean and covariance. In the present context, it allows one to summarise density dynamics using the method of moments (MM; (Rodriguez and Tuckwell, 1998; Rodriguez, 1996)). This entails replacing the system of stochastic differential equations with a system of deterministic equations (ODE) representing the dynamics of the means, variances, and covariance of the state variables, i.e., the first and second-order moments of the population density. This is precisely what we have done; namely, derive the ODE for the sufficient statistics of a Gaussian population density, given any set of Fokker-Planck equations that are coupled by phase-functions specifying mean-field effects or effective connectivity.

In related work, Hasegawa has proposed a semi-analytical mean-field approximation, in which the equations of motion for moments were derived for FitzHugh-Nagumo (FN) and Hodgkin-Huxley (HH) ensembles (Hasegawa, 2003a; Hasegawa, 2003b). Later he proposed an augmented moment method (AMM; (Hasegawa, 2004)), which relaxes the Gaussian or Laplace approximation (Hasegawa, 2006, 2007). In (Deco and Marti, 2007), the MM was extended to cover bimodal densities on the state variables; such that a reduced system of deterministic ODEs could be derived to characterise regimes of multistability. The ODEs in Figure 5.1 pertain to Morris-Lecar-like

neurons and will form the basis of dynamical causal models of empirical EEG data in the next Chapter.

5.7 Conclusion

We have derived a generic mean-field treatment of neuronal dynamics, which is based on a Laplace approximation to the ensemble density and is formulated in terms of equations of motion for the sufficient statistics of the ensemble density. We saw how this approach reduces to a neural-mass model when the second-order statistics (*i.e.*, variance) of neuronal states is ignored. In the next Chapter, we will use the Laplace and neural-mass approximations presented here as generative models of electrophysiological responses to sensory input. This Chapter will use Bayesian model comparison to compare both models and establish whether empirical responses contain evidence for a role of the variance in shaping population dynamics. This framework allow one to adjudicate between models that include the high-order statistics of neuronal states in predicting EEG time series and may also be important in the context of EEG-fMRI fusion; where power (second-order statistics) in electrical dynamics may be an important predictor of BOLD signals.

CHAPTER 6

A DCM STUDY OF MEAN-FIELD AND NEURAL-MASS MODELS OF NEURONAL DYNAMICS

In the previous Chapter, we presented a mean-field model of neuronal dynamics as observed with magneto and electroencephalography. Unlike neural-mass models, which consider only the mean activity of neuronal populations, mean-field models track both the mean and dispersion of population activity. This can be potentially important, if the mean affects the dispersion or *vice versa*. The mean-field model presented in the previous Chapter forms the basis of a dynamic causal model of observed electromagnetic signals below. In this Chapter, we compare mean-field and neural-mass models of electrophysiological responses using Bayesian model comparison. We used dynamical causal modelling to ask whether there is any evidence for a coupling between the mean and dispersion in observed electromagnetic responses. In particular, we used Bayesian model comparison to compare homologous mean-field and neural-mass models; and test whether empirical responses support a role for population variance in shaping neuronal dynamics. We addressed this question to mismatch negativity (MMN) and somatosensory evoked potential (SSEP) data; as representative examples of evoked responses with relatively slow and fast dynamics respectively. Our main conclusion is that neural-mass models appear quite sufficient for cognitive paradigms. However, there is clear evidence for an effect of dispersion at the high levels of depolarisation evoked in SEP paradigms. This suggests that (i) the dispersion of neuronal states within populations generating evoked brain signals can be manifest in observed brain signals and that (ii) the evidence for their effects can be accessed with dynamic causal model comparison.

6.1 Introduction

Neuronal responses are generated by the activity of coupled neuronal populations, as they respond to sensorimotor or cognitive perturbations. Models of these dynamics

allow one to ask questions about how observed data are generated. Neural-mass models (NMMs) have been used in this role for many years (David and Friston, 2003; Freeman, 1975; Jansen and Rit, 1995; Lopes da Silva et al., 1976; Nunez, 1974; Valdes et al., 1999; Wilson and Cowan, 1972). NMMs are economic models of the mean activity (e.g., firing rate or membrane potential) of neuronal populations and have been used to emulate a wide range of brain rhythms and dynamics (Amari, 1972; Deco et al., 2008; Frank et al., 2001; Haskell et al., 2001; Knight, 1972a, b; Nykamp and Tranchina, 2000; Omurtag et al., 2000; Robinson et al., 2005; Rodrigues, 2006; Sompolinsky and Zippelius, 1982).

In Chapter 5, we formulated neural-mass models, currently used as generative models in dynamic causal modelling, as a limiting case of mean-field models, in which the variance of the activity in any one neuronal population was fixed. Unlike neural-mass models, mean-field models consider the full density on the states of modelled populations including the variance or dispersion. We derived a generic mean-field treatment of neuronal populations or ensembles, based on a Laplace approximation to the population or ensemble density. This treatment was formulated in terms of equations of motion for the sufficient statistics of the ensemble density. Because a Gaussian density can be specified in terms of its first two moments, the ensuing scheme is formally identical to the second-moment method described by (Rodriguez, 1996). This reduces to a neural-mass model when the second-order statistics (*i.e.*, variance) of neuronal states is assumed to be constant. The key behaviour we were interested in was the coupling between the mean and variance of the mean-field Laplace approximation, which is lost in the neural-mass approximations. Here, we use the mean-field density dynamics as the basis of a dynamic causal model (DCM) of observed data. The resulting framework allowed us to adjudicate between models which include (or not) the high-order statistics of neuronal states when predicting EEG/MEG time series.

This Chapter comprises two sections. In the first, we summarize the DCM used here, in terms of the prior densities on the parameters of the mean-field neuronal model of the previous chapter and a mapping from hidden neuronal states to measurement space. In the second section, we use two EEG data sets and Bayesian model selection

(BMS) to assess the relative evidence for neural-mass and mean-field models. In addition, we establish the face-validity of neural-mass DCMs and their mean-field generalisations using synthetic data, generated using the conditional estimates of the network parameters, for each of the empirical examples.

6.2 Theory

Neural-mass and field models can reproduce neuronal dynamics reminiscent of observed evoked responses. However, to emulate more complex dynamics we may need to take into account the high-order statistics of ensemble dynamics. In Chapter 5 we derived a generic mean-field treatment of neuronal dynamics, based on a Laplace approximation to the ensemble density. This model is formulated in terms of equations of motion for the moments of the ensemble density, reducing to a NMM when the second-order moment (variance) is ignored. The most interesting behaviour in these mean-field models arises from the coupling between the mean and variance of ensemble activity, which is ignored in neural-mass approximations. Here, we will use the Laplace and neural-mass approximations in DCMs of electrophysiological responses to sensory input. See section 5.2 for a review on modelling neuronal dynamics with mean-field models and see section 5.3.4 for its application to a conductance-based model.

Dynamic Causal Modelling (DCM) provides a generative model for M/EEG responses (see Chapter 2; (David et al., 2006a; Kiebel et al., 2008)). The idea is that M/EEG data are the response of a dynamic system to experimental inputs, which are processed by a network of interacting neuronal sources. Here every source contains different neuronal populations (Figure 5.2), each described by a NMM (Equation 5.19) or a MFM (Equation 5.14). Each population has its own (intrinsic) dynamics governed by the neural-mass or the mean-field equations above, but also receives extrinsic input, either directly as sensory input or from other sources. The dynamics of these sources are specified fully by a set of first-order differential equations that are formally related to other neural-mass and mean-field models of M/EEG (e.g., (Breakspear et al., 2006; Rodrigues, 2006)).

The neuronal part of the DCMs in this Chapter was based on the mean-field model of the previous chapter (Equations 5.19 and 5.14). Table 6.1 lists the priors for the free MFM parameters and the values we used for its fixed parameters.

Free parameters	
Extrinsic coupling parameters	$\gamma_{ij}^{E} = \begin{cases} \frac{1}{2}e^{\theta} & forward \\ \frac{1}{4}e^{\theta} & backward \\ \frac{1}{4}e^{\theta} & lateral \end{cases} \quad p(\theta) = N(0,1)$
Intrinsic coupling parameters	$\gamma_{ij}^{E} = e^{\theta}\begin{bmatrix} 0 & 0 & \frac{1}{2} \\ 0 & 0 & 1 \\ \frac{1}{2} & 0 & 0 \end{bmatrix} \quad \gamma_{ij}^{I} = e^{\theta}\begin{bmatrix} 0 & \frac{1}{4} & 0 \\ 0 & 0 & 1 \\ 0 & 1 & 0 \end{bmatrix} \quad p(\theta) = N(0, \frac{1}{32})$
Capacitance	$C = 8e^{\theta} \quad p(\theta) = N(0, \frac{1}{32})$
Time constants	$\tau_e = \frac{1}{\kappa_e} = 4e^{\theta} ms \quad \tau_i = \frac{1}{\kappa_i} = 16ms \quad p(\theta) = N(0, \frac{1}{32})$
Diffusion tensor	$D(\omega) = e^{\theta}\begin{bmatrix} \frac{1}{8} & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \quad p(\theta) = N(0, \frac{1}{64})$
Conduction delays	$\Delta_{ij} = \begin{cases} 2e^{\theta} ms & intrinsic \\ 16e^{\theta} ms & extrinsic \end{cases} \quad p(\theta) = N(0, \frac{1}{128})$
Stimulus function parameters	$u(t) = \exp\left(-\frac{(t - \tau_u e^{\theta_{latency}})^2}{\tau_u e^{\theta_{duration}}}\right) : \begin{cases} \tau_u = 128ms : \text{MMN} \\ \tau_u = 16 \ \& \ 32ms : \text{SEP} \end{cases} \quad p(\theta_i) = N(0, \frac{1}{16})$
Fixed parameters	
Potentials	$V_L = -70\mu V \quad V_E = 60\mu V \quad V_I = -90\mu V_1 \quad V_R = -40\mu V$
Conductance	$g_L = 1$

Table 6.1: Prior densities of the MFM parameters.

As described in Chapter 2, to complete the specification of the DCM we need to specify how the hidden neuronal states map to observed responses. We assumed that the depolarization of pyramidal cell populations gives rise to observed M/EEG data,

which are expressed in the sensors through a conventional lead-field. The full spatiotemporal model takes the form of a nonlinear state-space model with hidden states modelling (unobserved) neuronal dynamics, while the observation (lead-field) equation is instantaneous and linear in the states. The ensuing DCM is specified in terms of its state-equation (Equation 5.19 or 5.14) and an observer or output equation

$$h(\theta) = L(\theta^L)\mu \tag{6.1}$$

where μ are the means above, $h(\theta)$ is the predicted signal and $\theta \supset \theta^L, \gamma_{ij}^k, \kappa_1, \kappa_E, w\ldots$ are unknown quantities that parameterize the state and observer equations. The parameters also control any unknown attributes of the stimulus function encoding exogenous input; we use a Gaussian density function parameterised by its onset and dispersion. We assume the MEG or EEG signal is a linear mixture of depolarisations in the pyramidal populations; where the columns of $L(\theta^L)$ are conventional lead-fields, which account for passive conduction of the electromagnetic field from the sources to the sensors. The parameters of the lead-field, θ^L encode the location and orientation of the underlying sources.

The predicted signal $h(\theta)$ corresponds to a generalized convolution of exogenous inputs (i.e., experimental stimulus functions). Under Gaussian assumptions about measurement noise, this generalized convolution gives a likelihood model for observed EEG or MEG data *y*

$$\begin{aligned} y &= vec\left(h(\theta) + X\theta^X\right) + \varepsilon \Rightarrow \\ p\left(y|\theta,\lambda\right) &= N\left(vec\left(h(\theta) + X\theta^X\right), diag\left(\lambda\right) \otimes V\right) \end{aligned} \tag{6.2}$$

Noise, ε, is assumed to be zero-mean Gaussian and independent over channels, where λ is a vector of unknown channel-specific error variances and V represents a temporal autocorrelation matrix. Low-frequency noise or drift components are modelled by confounding variables in the columns of the matrix, X (this was simply a constant term in this paper). For computational expediency, we reduce the dimensionality of the sensor data, while retaining the maximum amount of information. This is assured by projecting the data onto a subspace defined by its

principal modes; computed using singular value decomposition. The likelihood in Equation 6.2 and the priors in Table 6.1 complete the DCM specification and allow it to be inverted for any given data in the usual way (see Appendix B).

As we saw in previous Chapters, a DCM is fitted to data by tuning the free parameters to minimize the discrepancy between predicted and observed MEG/EEG time series, under complexity constraints. In addition to minimizing prediction error, the parameters are constrained by a prior specification of the range they are likely to lie in (Friston et al. 2003). These constraints, which take the form of a prior density - $p(\vartheta)$, are combined with the likelihood, $p(y|\vartheta)$, to form a posterior density $p(\vartheta|y) \propto p(y|\vartheta)p(\vartheta)$ according to Bayes' rule. The priors $p(\vartheta)$ are usually specified under log-normal assumptions to impose positivity constraints; and are therefore specified by the prior mean and variance of log-parameters. Table 6.1 lists the priors for the free parameters of the neuronal model and the values we used for its fixed parameters.

The log-evidence is an important quantity because it allows one to compare different models, (Penny et al. 2004). The most likely model is the one with the largest log-evidence. Model comparison rests on the likelihood ratio (i.e., Bayes-Factor) of the evidence or relative log-evidence for two models. Strong evidence in favour of one model typically requires the difference in log-evidence to be three or more (Penny et al. 2004). Under flat priors on models this corresponds to a conditional confidence that the winning model is $\exp(3) \approx 20$ times more likely than the alternative. This indicates that the data provide 'strong' (10:1 to 30:1) evidence in favour of one model over the other. See http://en.wikipedia.org/wiki/Bayes_factor for the range of Bayes factors indicating 'very strong' (30:1 to 100:1) and 'decisive' (more than 100:1) evidence for a model. In the next section, we will use the free-energy bound on log-evidence to compare the different models elaborated above.

6.3 Simulations and empirical results

Our key question was: can we find evidence for coupling between the mean and dispersion of neuronal states in empirical data? However, we anticipated that the answer would be context sensitive; in the sense that some evoked responses may induce large fluctuations in dispersion, whereas others may not. This context-sensitivity can be seen from the form of Equation 5.14, where changes in the dispersion of neuronal states depend upon the systems Jacobian $\partial_x f^{(i)}$ or rate of change of flow with state. The Jacobian depends on depolarisation and conductance (Equation 5.18), which depends on presynaptic input $\varsigma_k^{(i)}$. This implies that we would expect to see large fluctuations in dispersion and the ensuing effect on the mean under high levels of extrinsic presynaptic input. We therefore chose to perform our model comparison using two sorts of evoked responses. The first used a traditional 'cognitive' paradigm (a mismatch negativity paradigm) in which auditory stimuli can be regarded as delivering low amplitude physiological inputs to cortical sources. In contrast, the second paradigm was a somatosensory evoked potential (SEP) paradigm; in which neuronal sources are excited with a non-physiological electrical stimulus, eliciting transient but high amplitude presynaptic inputs. We predicted that if there was any evidence for the mean-field model, relative to the neural-mass model, then we would be more likely to see it in the SEP paradigm, relative to the mismatch negativity paradigm. In what follows, we describe these paradigms and the results of our model comparisons.

6.3.1 Mismatch Negativity Paradigm

In this section, we analyze data from a multi-subject mismatch negativity (MMN) (Garrido et al., 2007b). The MMN is the differential response to an unexpected (rare or oddball) auditory stimulus relative to an expected (standard) stimulus. The MMN has been studied extensively and is regarded as a marker for error detection, caused by a deviation from a learned regularity, or familiar auditory context. According to (Näätänen et al., 2001) the MMN is caused by two underlying functional processes, a sensory memory mechanism and an automatic attention-switching process that might

engage frontal generators (Giard et al., 1990). It has been shown that the temporal and frontal MMN sources have distinct behaviours over time (Rinne et al., 2000) and that these sources interact with each other (Jemel et al., 2002). Thus the MMN could be generated by a temporofrontal network (Doeller et al., 2003; Escera et al., 2003; Opitz et al., 2002), as revealed by M/EEG and fMRI studies. In a predictive coding framework, these findings can also be framed as adaptation and experience-dependent plasticity in an auditory network (Friston, 2005; Garrido et al., 2007a; Garrido et al., 2007b; Jääskeläinen et al., 2004).

Using DCM, we modelled the MMN generators with a temporofrontal network comprising bilateral sources over the primary and secondary auditory and frontal cortex. Following Garrido et al. (2007a), we used a five-source network with forward and backward extrinsic (between-source) connections. Exogenous or auditory input (modelled with $u(t) \in \Re$, a parameterised bump function of time; see Table 6.1) enters via subcortical structures into two bilateral sources in posterior auditory cortex (**lA1** and **rA1**). These have forward connections to two bilateral sources in anterior auditory cortex; i.e., superior temporal gyri (**lSTG** and **rSTG**). These sources are laterally and reciprocally connected via the corpus callosum. The fifth source is located in the right inferior frontal gyrus (**rIFG**) and is connected to the **rSTG** with reciprocal unilateral connections. Using these sources and prior knowledge about the functional anatomy cited above, we specified the DCM network in Figure 6.1. Here, we were interested in comparing the NMM and MFM formulations of this network, in terms of their negative free-energy. To simplify the analysis, we modelled only the responses evoked by standard stimuli (from 0 ms to 256 ms).

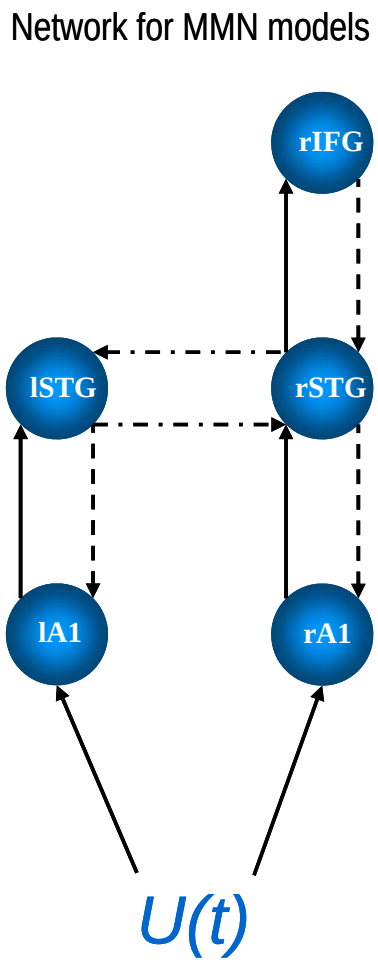


Figure 6.1: DCM network used for the mismatch negativity paradigm for both models; NMM and MFM. Forward connections (full lines), backward connections (dash lines), and lateral connections (dash-dot lines) couple sources. A1: primary auditory cortex, STG: superior temporal gyrus, IFG: inferior temporal gyrus. l and r – left and right brain hemispheres respectively. *U(t)* is the auditory input stimuli driving the network.

6.3.1.1 *Empirical results*

Two DCMs (NMM and MFM variants) were inverted for all twelve subjects and compared using their log-evidence. Figure 6.2 shows the differences in log-evidences for each subject. For all but one subject, there was decisive evidence for the NMM over the MFM. The log-evidence at the group level (>100), pooled over all subjects

(given the data are conditionally independent over subjects) was similarly decisive. Although the relative log-evidence is quantitatively meaningful in its own right, one can also treat it as a log-odds ratio and use its distribution over subjects to compute a classical *p*-value (Stephan et al 2009). In this instance, a one-sample *T*-test was extremely significant ($T = 3.58$, d.f. $= 11$, $p = 0.002$). This means that we can reject the null hypothesis that the data are explained equally well by neural-mass and mean-field formulations of the same neuronal model.

These results suggest the NMM is a better model for explaining evoked auditory responses. Note that the complexity of the NMM and MFM are the same; the MFM has more states but does not have more unknown parameters. This may seem counterintuitive because the dispersion in the MFM may appear to make it more complicated. However, the dispersion is a sufficient static of a density on hidden states and is not itself subject to random effects. This means, given the model parameters, it is a deterministic quantity and does not add to model complexity (i.e., it is specified by the same parameters as the neural mass model). This is important because the results in Figure 6.2 are remarkably consistent over subjects and cannot be explained by differences in model complexity. In short, the NMM provides a better prediction of the observed responses than the MFM, in this paradigm. Furthermore, the differences between the NMM and MFM predictions are fairly subtle. This suggests that the population variance is actually quite stable over peristimulus time, because the model selection clearly favours the predictions from the neural-mass model.

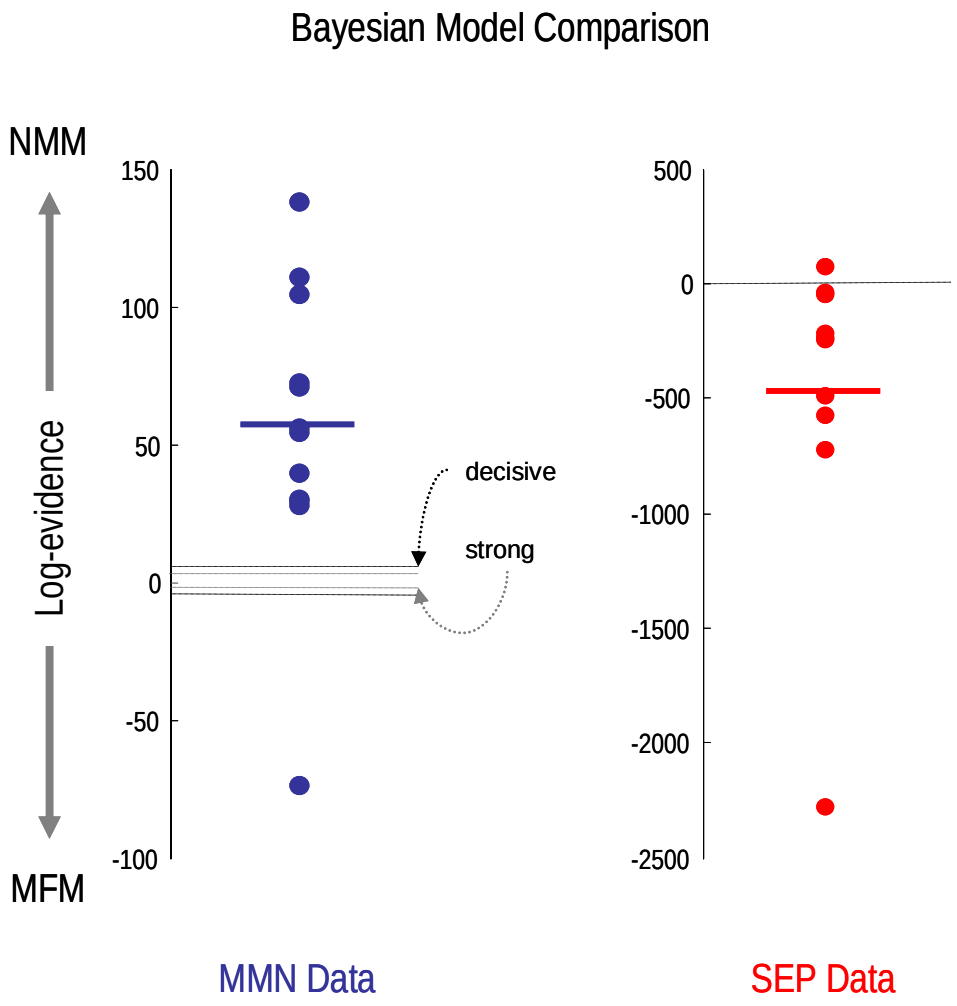


Figure 6.2: Bayesian model comparisons for NMM in relation to MFM. Right: Relative log-evidence for the NMM for each subject using the network in Figure 2. The NMM log-evidences are consistently better than the MFM log-evidences, with a pooled difference >100 over subjects. Left: the same results for the SEP data; the group log-evidence difference was >100 in favour of the MFM. The solid lines indicate the mean over subjects.

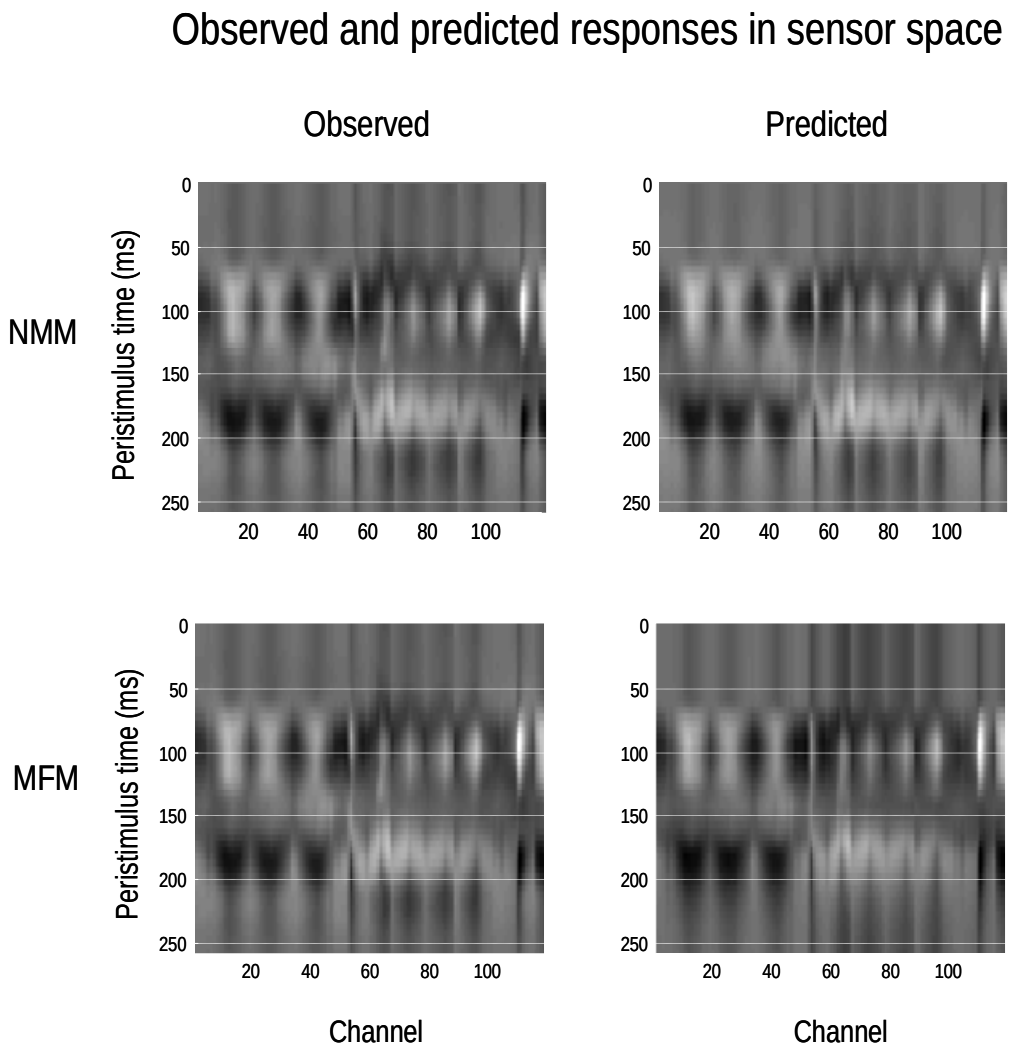


Figure 6.3: **Upper panels:** Observed (left) and predicted (right) evoked responses over 128 channels and peristimulus time shown in image formation (grey scale normalised to the maximum of each image). These came from the NMM-DCM of the first subject. **Lower panels:** MFM predictions for the same subject. We show the observed response twice because they are adjusted for the confounding DC or constant term in our model (see Equation 2.6). This adjustment renders the observed data slightly different, depending on the model fit.

6.3.1.2 *Simulations*

We next performed some comparative evaluations and validations of DCM using neural-mass and mean-field models, using synthetic data based on the empirical results above. These are presented to show that the empirical model comparison above is sufficiently sensitive to disambiguate between neural-mass and mean field variants of the same model. After generating data from a known model, we used model comparison to ask whether one can recover the correct model over its alternative. We integrated the NMM and MFM with known (true) model parameters derived from the real data above (the conditional means from a DCM of the grand average over subjects) and added random measurement noise with a standard deviation of 10% of the peak response in channel space. This was roughly the amplitude of noise in the real data. Finally, we used the synthetic data generated by both models to invert the neural-mass and mean-field DCMs. Table 6.2 lists the resulting log-evidences. Each column contains the log-evidences for each data-set. The maximum values are found on the diagonal; i.e., the true model had the greatest evidence and that the relative evidence for the correct model was 'decisive'. These results confirm that these models can be disambiguated using DCM, under empirically realistic levels of noise.

Models	***Synthetic Data***	
	NMM	**MFM**
NMM	**-662.7**	-952.9
MFM	-844.2	**-665.5**

Table 6.2: Log-evidences for neural-mass (NMM) and mean-field (MFM) models using synthetic data generated by a five-source MMN model (see Figure 6.1) using NMM and MFM formulations. The diagonal values show higher log-evidences for the true model.

6.3.2 Somatosensory Evoked Potential Paradigm

To explore the context sensitivity of these results, we analyzed data from a study of paired associative stimulation (PAS), Litvak et al. (2007), which involves repetitive magnetic cortical stimulation timed to interact with median nerve stimulation-induced peripheral signals from the hand. The PAS paradigm has been shown to induce long-lasting changes in somatosensory evoked potentials (Wolters et al., 2005) as measured by single-channel recordings overlying somatosensory cortex. The SEP generators evoked by compound nerve stimulation have been studied extensively with both invasive and non-invasive methods in humans and in animal models (Allison et al., 1991). Litvak, et al. (2007) characterised the topographical distribution of PAS-induced excitability changes as a function of the timing and composition of afferent somatosensory stimulation, with respect to a transcranial magnetic stimulation (TMS). The temporal response pattern of the SEP comprises a P14 component generated subcortically and then a N20–P30 complex from the sensorimotor cortex, which is followed by a P25–N35 complex (Allison et al. 1991). The remainder of the SEP can be explained by a source originating from the hand representation in S1 (Litvak et al., 2007).

We chose these data as examples of fast sensory transients that might engage a more circumscribed network than the auditory stimuli in the MMN paradigm above. We anticipated that the density dynamics of neuronal populations that were stimulated electromagnetically, may disclose the effects of dispersion (see above). We analysed the SEP data from eleven subjects following median nerve stimulation (i.e., in the absence of transcranial magnetic stimulation) as above. The network architecture was based on previous reports (Buchner et al., 1995; Ravazzani et al., 1995, Litvak et al., 2007): we modelled the somatosensory system with three sources, each comprising three neuronal populations. In this model (see Figure 6.4, Litvak et al., 2007 and Marreiros et al 2008) exogenous input was delivered to the brainstem source (**BS**), which accounts for early responses in the medulla. The input was a mixture of two parameterised bump functions with prior latencies based on known conduction delays (see Table 6.1). This region connects to two sources **SI** and **SII** in Brodmann area 3 (Marreiros et al., 2008). We inverted the resulting DCMs using the sensor data from 4

ms to 64 ms, following stimulation. We report the results form the first ten subjects because the DCM inversion failed to converge for the last subject.

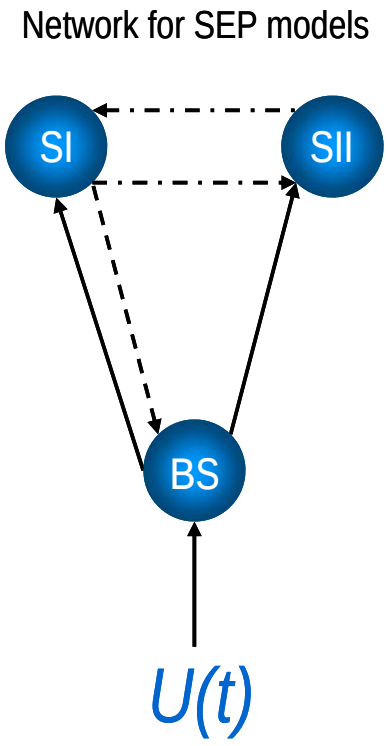


Figure 6.4: DCM network used for the SEP paradigm and both NMM and MFM-based DCMs. Forward connections (full lines), backward connections (dash lines) and lateral connections (dash-dot lines) connect the sources. BS: brainstem source, SI and SII: two somatosensory sources on Brodmann area 3b. *U(t)* is the median nerve input stimuli driving the network.

6.3.2.1 *Empirical results*

Figure 6.2 (right) shows the log-evidence differences. In stark contrast to the MMN results, there was 'decisive' evidence in all but one subject for the MFM over the NMM. Moreover, the large group difference in log-evidence of (>100) favours the MFM model; i.e., if we had to account for the data from all subjects with the same model, then the evidence for the MFM was decisive. The classical *p*-value was similarly significant (T = 2.19, d.f. = 9, p = 0.028) but less significant than in the

MMN analyses due to larger inter-subject variability. These results indicate that the MFM is, in this instance, a demonstrably better explanation for somatosensory evoked potentials. It is important to appreciate that exactly the same model was used for the MMN and SEP data, including the prior density on the free parameters (with the exception of the exogenous input). However, the results of model comparison are, as anticipated, completely the opposite and remarkably consistent over subjects for both paradigms.

Anecdotally, the superior performance of the mean-field model seemed to be its ability to fit both the early N20-P30 complex and later waveform components of the SEP (although no model was able to fit the P14 components convincingly). This contrasts with the neural-mass model that was unable to reproduce the fast initial transients but was able to model the slower components that followed. Phenomenologically, this means the dispersion of neuronal states in the MFM confers a greater range on the time constants of population dynamics, which allows the MFM to reproduce fast, large-amplitude responses in, we presume, relatively circumscribed neuronal populations responding synchronously to extrinsic afferents.

6.3.2.2 *Simulations*

To ensure the model comparison retained its sensitivity, in this SEP setting, we again generated synthetic data using the conditional means of the parameters estimated from the empirical data. We used a NMM and a MFM for generation and inversion and evaluated both combinations to ensure that model selection identified the correct model. For the integration of the forward models, we used the conditional means of parameters from an analysis of the grand-average data across subjects. We added random noise to these synthetic data, with a standard deviation that was 5% of the peak response in sensor space. We used the three source model above (Litvak et al., 2007; Marreiros et al., 2008a) to generate and model the data. Table 6.3 presents the log-evidences for each of the four inversions. The highest evidences were obtained for the models that were used to generate the synthetic data: these correspond to the

diagonal entries. Again, the results conform that model comparison can identify the correct model of these somatosensory responses.

Models	*Synthetic Data*	
	NMM	**MFM**
NMM	**-185.0**	-175.8
MFM	-369.6	**-102.1**

Table 6.3: Log-evidences for neural-mass (NMM) and mean-field (MFM) models using synthetic data generated by a three-source SEP model (see Figure 6.4) using NMM and MFM formulations. The diagonal values show higher log-evidences for the true model.

6.4 A quantitative illustration of density dynamics

Figure 6.5 (upper left panel), shows the sufficient statistics of population activity for source in the first SEP subject. These are the *mean* and *covariance* of neuronal states, in source space. These are obtained by integrating the ensemble dynamics in Equation 5.14, using the equations of motion in Equation 5.21 (and Figure 5.2) and the conditional parameter estimates. Generally, when the mean depolarization increases, the covariance decreases, getting close to zero when the mean approaches its maximum. This is seen here at about 30 ms. This concentration of neuronal states augments the decay of mean depolarisation (see Equation 5.14). Note that at around 20 ms the N20 is modelled by polyphasic dynamics in the mean depolarisation that rest on a coupling with dispersion. It is this coupling and ensuing dynamics that are missing in the neural mass model. In the lower panel, we see the conditional response estimate, in sensor space, in terms of the observed (dotted lines) and predicted (solid lines) time-series for all modes. These results are representative of DCM prediction accuracy.

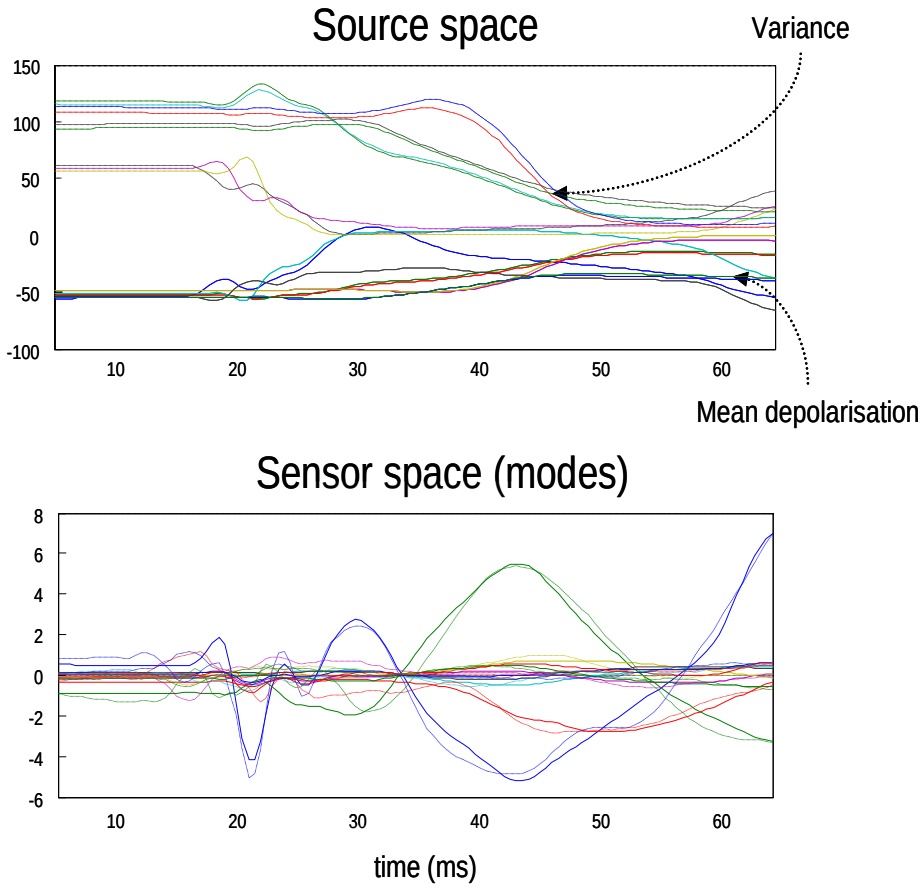


Figure 6.5: Standard DCM output for SEP data (right) for 64 ms peri-stimulus time and MMN data (left) for 256 ms peri-stimulus time. Upper panels: Conditional estimates of the mean and covariance of neuronal states, in source space (coloured lines correspond to different neuronal subpopulations). Lower panels: Conditional estimates of responses, in sensor space (coloured lines correspond to different spatial modes; solid line: predicted; dotted line: observed).

6.5 Discussion

We have introduced a mean-field DCM for M/EEG data, which approximates the full density dynamics of population activity in neuronal sources of observed electromagnetic responses with a Gaussian density. This work was motivated by the observation that neural-mass models, which consider only the first moment of the density of each neuronal population, can be seen as a limiting case of mean-field

models (Chapter 5). The mean-field model used physiological plausible priors with the hope of creating a reasonably realistic conductance-based model. We have shown, using model inversion and simulations that one can disambiguate between MFM and NMM models (Tables 6.2 and 6.3) and found that the NMM was the best model for explaining the MMN data. In contrast, we found that the MFM was a better model of the SEP data (Figure 6.2), in the vast majority of subjects and at the group level. We deliberately chose these distinct data-sets in the hope of disclosing this dissociation between neural-mass and mean-field models:

This difference in performance between the two models on the two data-sets lies in the difference between Equations 5.14 and 5.19. Our results suggest that the MFM captures the faster SEP population dynamics better than the NMM. This may be because the SEP paradigm evokes larger presynaptic inputs to small circumscribed neuronal populations as compared to the MMN and related cognitive paradigms. It is this input that induces changes in conductance which synchronise hidden states and cause a subsequent suppression of mean activity. It can be seen from Equation 5.14 that changes in covariance depend on the derivative of flow. Equation 5.18 shows that this depends on depolarisation and conductance. In support of this, the MFM solutions for the SEP data do indeed show a reciprocal coupling between mean depolarisation and variance (see Figure 6.5). Having said this, the appropriateness of a model for any particular data or paradigm data cannot necessarily be deduced analytically. The aim of this paper is to show that questions about density dynamics of this sort can be answered using Bayesian model comparison. Future studies with NMM and MFM may provide heuristics about the relative utility of these models. In particular, it will be interesting to use MFMs when more complex dynamics are induced by extreme perturbations from steady-state dynamics (e.g., transcranial magnetic stimulation).

Most DCMs in the literature are deterministic; in that they allow for observation noise on the sensors but do not consider random fluctuations on hidden states. Here, the hidden states in mean-field DCMs are sufficient statistics of a density, which accommodates random fluctuations on neuronal states. This is important because it means we can model systems specified in terms of stochastic differential equations (cf. Equation 5.21) with ordinary differential equations (Equation 5.14) through the

Fokker-Planck formalism. A potential application of this approach, beyond finessing prediction of EEG signals, could be in the context of EEG-fMRI fusion; where the second-order statistics of neuronal activity (c.f. power) may be an important predictor of BOLD signals.

It is important to appreciate that the scheme described in this paper is not tied to any particular model of neuronal dynamics. The same comparisons presented above could be repeated easily, using any model of neuronal dynamics that are entailed by their equations of motion. Indeed, we anticipate that people will want to compare neural-mass and mean-field implementations of different models (see software note). The only constraint on these comparisons is that the equations of motion should be nonlinear in the neuronal states. This is because linear models preclude a coupling of first and second-order moments and render the behaviour of the neural mass and mean field formulation identical.

6.6 Conclusion

We have shown that it is possible to implement a mean-field DCM, which considers the mean and variance of neuronal population activity. The modulation of second-order statistics may be a useful extension of DCM for evoked responses, as measured with magneto- and electroencephalography. Critically, the role of higher moments can be assessed empirically in a Bayesian model comparison framework. In this initial work, we conclude that, although conventional neural-mass models are probably sufficient for most applications, it is easy to find strong evidence for coupling among the moments of neuronal ensemble densities in observed EEG data.

CHAPTER 7

GENERAL DISCUSSION AND CONCLUSION

7.1 Synopsis

We started off by building a two-state DCM for fMRI which allows inferences that can be meaningfully linked to specific neurotransmitter systems and permits the modelling of both extrinsic and intrinsic connections. Our results indicate that it is possible to estimate area-intrinsic connection strengths using fMRI within network models. With real data, using Bayesian model selection, we found that the two-state DCM is a better model than the standard single state DCM. This demonstrated the potential of adopting generative models that are informed by anatomical and physiological principles.

In the second chapter, we saw that the sigmoid activation function currently used in neural mass models can be interpreted as a cumulative density function on depolarisation. We then looked at how the dynamics of a population can change profoundly when the variance (slope-parameter) changes. In particular, we examined how the input-output properties of populations depend on the sigmoid, in terms of first (driving) and second (modulatory) order convolution kernels and corresponding transfer functions. Using real data we showed that the population variance can be quite substantial: Using DCM, we quantified the population variance in relation to the evolution of mean activity of neural-masses. The quantitative results of this analysis suggested that only a small proportion of neurons are actually firing at any time, even during the peak of evoked responses.

The insights from the previous studies motivated a more general model of population dynamics. Thus, we derived a generic mean-field treatment of neuronal dynamics, based on a Laplace approximation to the ensemble density and formulated in terms of equations of motion for the sufficient statistics of the ensemble density. We saw how this approach reduces to a neural-mass model when the second-order statistics (*i.e.*, variance) of neuronal states is ignored.

Subsequently, we implemented a mean-field DCM for ERPs based on a conductance-voltage microcircuit, which considers both the mean and variance of neuronal population activity. We saw that the modulation of second-order statistics may be a useful extension of DCM for evoked responses, as measured with MEG and EEG. Critically, the role of higher moments can now be assessed empirically in a BMS framework. In this initial work, we found strong evidence for coupling among the first moments (i.e., mean and variance) of neuronal ensemble densities in observed EEG data.

7.2 General summary

The aim of the work described in the first result section, **Chapter 3**, was to endow dynamic causal models (DCM) for fMRI time series with a greater biological realism. We have described a new DCM for fMRI, which has two states per region instead of one. With the two-state DCM, it is possible to relax shrinkage priors used to guarantee stability in single-state DCMs. Moreover, we can model both extrinsic and intrinsic connections, as well as enforce positivity constraints on the extrinsic connections. Using synthetic data, we have shown that the two-state model has internal consistency. We have also applied the model to real data, explicitly modelling intrinsic connections. Using model comparison, we found that the two-state model is better than the single-state model and that it is possible to disambiguate between subtle changes in coupling. These results suggest that the parameterization of the standard single-state DCM is possibly too constrained. With a two-state model, the data can be better explained by richer dynamics (and more parameters) at the neuronal level. This study demonstrated the potential of adopting generative models for fMRI time-series that are informed by anatomical and physiological principles.

Having compared DCMs with one or two neuronal states per brain region for fMRI data, we turned to DCM for EEG and MEG data. Specifically, we evaluated DCMs based on density-dynamics. To ensure sufficient temporal precision in the data, we moved from haemodynamic responses to electrophysiological responses such as the

ERP measured with EEG or MEG. In this context, neural-mass DCMs generally have a fixed variance because they assume a fixed-form for the sigmoid activation function. NMMs are obliged to make this assumption because their state variables allow only changes in mean states, not changes in variance or higher-order statistics of neuronal activity. Is this assumption sensible?

In **Chapter 4** our focus was on how the sigmoid activation function, linking mean population depolarization to expected firing rate, can be understood in terms of the variance or dispersion of neuronal states. We showed that the slope-parameter ρ models formally the effects of variance (to a first approximation) on neuronal interactions. Specifically, we saw that the sigmoid function can be interpreted as a cumulative density function on depolarisation, within a population. We looked at how the dynamics of a population can change profoundly when the variance (slope-parameter) changes. In particular, we examined how the input-output properties of populations depend on ρ, in terms of first (driving) and second (modulatory) order convolution kernels and corresponding transfer functions. We used real EEG data to show that population variance, in the depolarisation of neurons from somatosensory sources generating SEPs, can be quite substantial. Using DCM, we estimated the SEP parameter density controlling the shape of the sigmoid function. This allowed us to quantify the population variance in relation to the evolution of mean activity of neural-masses and provided anecdotal evidence for changes in variance over different time-windows of the data. The quantitative results of this analysis suggested that only a small proportion of neurons are actually firing at any time, even during the peak of evoked responses.

This Chapter motivated a more general model of population dynamics which compared DCMs based on density-dynamics with those based on neural-mass models. These models allowed us to ask if the variance of neuronal states in a population affects the mean (or *vice versa*) using the evidence or marginal likelihood of the data under different models. Moreover, we could see if observed responses are best explained by mean firing rates, or some mixture of the mean and higher-order moments. This would allow one to adjudicate between models that include high-order statistics of neuronal states in EEG time-series models.

In **Chapter 5**, we derived a generic mean-field treatment of neuronal dynamics, which is based on a Laplace approximation to the ensemble density and is formulated in terms of equations of motion for the sufficient statistics of the ensemble density. The high dimensionality and complexity of Fokker-Planck formalisms can be reduced with a mean-field approximation, which describes the evolution of separate ensembles coupled by mean-field effects. By parameterising the densities in terms of their sufficient statistics, the partial differential equations can be reduced to ordinary differential equations describing the evolution of its sufficient statistics or moments (Hasegawa, 2003a; Hasegawa, 2003b). In this way, we obtained a key equation, which formulates population dynamics, using only the flow, its gradient and curvature, at the mean state. The key behaviour we were interested in was the coupling between the mean and variance of the ensemble, which is lost in the neural-mass approximations. This enabled us to compare equivalent mean-field and neural-mass models of the same populations and evaluate, quantitatively, the contribution of population variance to shaping population dynamics. We compared the Laplace approximation with the neural-mass model to assess the role of the population covariance. Qualitative differences between MFM and NMM predictions were easy to demonstrate. The MFM showed a dynamical behaviour more similar than the NMM to the response obtained by integrating the stochastic ensemble dynamics. In particular, we saw that the MFM showed a bifurcation from fixed-point to a limit-cycle attractor, as sustained input levels were increased. The spectral response of the pyramidal population of the mean-field model analysis disclosed the presence of physiologically plausible oscillatory signals in the alpha and beta band and how their relative power changed with activation. Additionally, we presented an interesting example of the quantitative difference between MFM and NMM by reproducing nested oscillation behaviour. In short, the MFM appeared to represent richer and more complex dynamics. This may have important valuable applications in DCM of imaging studies (M/EEG, fMRI), where one tries to explain the coordinated activity of a large number of neurons.

In **Chapter 6**, we used the Laplace and neural-mass approximations of the previous Chapter as generative models of electrophysiological responses to sensory input. We introduced a mean-field DCM for M/EEG data, which considers the mean and variance of neuronal population activity. This work was motivated by the observation that neural-mass models, which consider only the first moment of the density of each neuronal population, can be seen as a limiting case of mean-field models, (Chapter 4).

We have shown, using model inversion and simulations that one can disambiguate between MFM and NMM models and found that the NMM was the best model for explaining MMN data. In contrast, we found that the MFM was the best model for explaining SEP data. Our results suggest that the MFM captures the faster SEP population dynamics better than the NMM. Future studies with NMM and MFM may provide heuristics about the relative utility of these models. In particular, it will be interesting to use MFMs when more complex dynamics are induced by extreme perturbations from steady-state dynamics (e.g., transcranial magnetic stimulation). The modulation of second-order statistics may be a useful extension of DCM for evoked responses, as measured with magneto- and electroencephalography. Critically, the role of higher moments can be assessed empirically in a Bayesian model comparison framework. In this initial work, we conclude that, although conventional neural-mass models are probably sufficient for various applications, it is easy to find strong evidence for coupling among the moments of neuronal ensemble densities in observed EEG data.

7.3 Future Directions

This section discusses potential extensions to DCM that may allow useful questions to be addressed.

7.3.1 *DCM for fMRI*

Currently, with the two-state DCM, we model excitatory (glutamatergic) and inhibitory (GABA-ergic) connections. As a natural extension, we can include further states per region, accounting for other neurotransmitter effects. Important examples

here would be adaptation phenomena and activity-dependent effects of the sort mediated by NMDA receptors. This is interesting because NMDA receptors are thought to be targeted preferentially by backward connections. This could be tested empirically using a suitable multi-state DCM based on an explicit neural-mass model. Another important point is that the haemodynamics in the current DCM are a function of the excitatory states only. The contributions to the BOLD signal from the inhibitory states are expressed indirectly, through dynamic interactions between the two states, at the neuronal level. One possible extension would be to model directly separate contributions of these two states, at the haemodynamic level. Hypotheses about the influence of excitatory and inhibitory populations on the BOLD signal could then be tested using model comparison.

Another extension is to generalize the interactions between the two subpopulations, i.e., to use nonlinear functions of the states in the DCM. Currently, this is purely linear in the states, but one could use sigmoidal functions. This would take our model into the class described by (Wilson and Cowan, 1973). In this way, one can construct more biologically constrained response functions and bring DCMs for fMRI closer to those being developed for EEG and MEG. Again, the question of whether fMRI data can inform such neural-mass models can be answered simply by model comparison.

Current DCMs do not account for noise on the states (i.e., random fluctuations in neuronal activity). There has already been much progress in the solution of stochastic differential equations entailed by stochastic DCMs, particularly in the context of neural mass models (see (Sotero et al., 2007; Valdes et al., 1999)). A number of methodological developments have improved and extended DCM for fMRI, e.g. Bayesian model selection amongst alternative DCMs (Penny et al., 2004a), precise sampling from predicted responses (Kiebel et al., 2007b), refined haemodynamic model (Stephan et al., 2007c) and a nonlinear DCM for fMRI (Stephan et al., 2008). These could all be combined with multistate DCMs for fMRI.

7.3.2 *DCM for ERP*

The hidden states in mean-field DCMs used here are sufficient statistics of a density, which accommodates random fluctuations on neuronal states. This is important because it means we can model systems specified in terms of stochastic differential equations with ordinary differential equations through the use of the Fokker-Planck formalism. A potential application of this approach, beyond finessing prediction of EEG signals, could be in the context of EEG-fMRI fusion; where the second-order statistics of neuronal activity (c.f. power) may be an important predictor of BOLD signals. Further development of M/EEG models and their fusion with other imaging modalities requires more complex models embodying useful constraints. The appropriateness of such models for any given data cannot necessarily be intuited, but can be assessed formally using Bayesian model comparison. Bayesian model comparison will probably become a ubiquitous tool in M/EEG (and fMRI).

A number of methodological developments have improved and extended DCM for ERP, a DCM for intrinsic connections (Kiebel et al., 2007a), a DCM for induced responses (Chen et al., 2008), a DCM of steady state responses (Moran et al., 2009). It can be expected that this trend will be considerably reinforced and accelerated during the next few years, fuelled by the need for mechanistic explanations of how cognition is mediated by neural systems and by the availability of more powerful modelling techniques.

7.3.3 *DCM and clinical applications*

The generic framework of DCM and the ongoing developments, will contribute to a more mechanistic understanding of brain function. Of particular interest will be the use of neural system models like DCM to (i) understand the mechanisms of drugs and (ii) to develop models that can serve as diagnostic tools for diseases linked to abnormalities of connectivity and synaptic plasticity. Concerning pharmacology, many drugs used in psychiatry and neurology change synaptic transmission and thus functional coupling between neurons. Therefore, their therapeutic effects cannot be

fully understood without models of drug-induced connectivity changes in particular neural systems.

The success of this approach will partially depend on developing models that include additional levels of biological detail while being parsimonious enough to ensure mathematical identifiability and physiological interpretability; see (Breakspear et al., 2003b; Harrison et al., 2005; Jirsa, 2004; Robinson et al., 2001) for examples that move in this direction. Another important goal is to explore the utility of models of effective connectivity as diagnostic tools (Stephan, 2004). This seems particularly attractive for neurological and psychiatric diseases whose phenotypes are often very heterogeneous and where a lack of focal brain pathologies points to abnormal connectivity (dysconnectivity) as the cause of the illness.

A major challenge will be to establish neural systems models which are sensitive enough that their connectivity parameters can be used reliably for diagnostic classification and treatment response prediction of individual patients. Ideally, such models should be used in conjunction with paradigms that are minimally dependent on patient compliance and are not confounded by factors like attention or performance. Given established validity and sufficient sensitivity and specificity of such a model, one could use it in analogy to biochemical tests in internal medicine, i.e. to compare a particular model parameter (or combinations thereof) against a reference distribution derived from a healthy population (Stephan, 2004).

APPENDICES

Appendix A

Psychophysiological interactions

(Büchel et al., 1996) discuss a series of increasingly high-order interaction terms in general linear models. These are introduced as new explanatory variables enabling SPM to estimate the magnitude and significance of nonlinear effects directly. A special example of this is a psychophysiological interaction (Friston et al., 1997), where the bilinear term represents an interaction between an input or psychological variable and a response or physiological variable y^i measured at the *i*-th brain region. Any linear model can be augmented to include a PPI

$$Y = [X \quad u \times y^i]\beta + \varepsilon . \tag{A.1}$$

The design matrix partition $X = [u, y^i, \ldots]$ normally contains the main effect of experimental input and regional response. The PPI is the Hadamard product $u \times y^i$ and is obtained by multiplying the input and response vectors element by element. Both the main-effects and interaction terms are included because the main effects have to be modelled to assess properly the additional explanatory power afforded by the bilinear or PPI term. PPI models provide important evidence for the interactions among distributed brain systems and enabled inferences about task-dependent plasticity using a relatively simple procedure.

Structural equation modelling

This model was developed explicitly with effective connectivity or path analysis in mind and rests on specifying constraints on the connectivity. There is no designed perturbation and the inputs are treated as unknown and stochastic. Furthermore, the inputs are often assumed to express themselves instantaneously such that, at the point of observation the change in states is zero. In the absence of bilinear effects we have

$$\begin{aligned} \dot{x} &= 0 \\ &= Ax + Cu \,. \\ x &= -A^{-1}Cu \end{aligned} \tag{A.2}$$

This is the regression equation used in SEM where $A = D - I$ and D contains the off-diagonal connections among regions. The key point here is that A is estimated by assuming $u(t)$ is some random innovation with known covariance. This is not really tenable for designed experiments when $u(t)$ represent carefully structured experimental inputs. Although SEM and related autoregressive techniques are useful for establishing dependence among regional responses, they are not surrogates for informed causal models based on the underlying dynamics of these responses.

Appendix B

Expectation Maximisation

This appendix describes **EM** for linear models using statistical mechanics (Neal and Hinton, 1998). We connect this formulation with classical methods and show the variational free energy is the same as the objective function maximised in restricted maximum likelihood (ReML).

The **EM** algorithm is ubiquitous in the sense that many estimation procedures can be formulated as such, from mixture models through to factor analysis. Its objective is to maximise the likelihood of observed data $p(y \mid \lambda)$, conditional on some hyperparameters, in the presence of unobserved variables or parameters θ. This is equivalent to maximising the log-likelihood

$$\begin{aligned} \ln p(y \mid \lambda) &= \ln \int p(\theta, \mid \lambda) d\theta \geq \\ F(q, \lambda) &= \int q(\theta) \ln p(\theta, y \mid \lambda) d\theta - \int q(\theta) \ln q(\theta) d\theta \end{aligned} \tag{B.1}$$

where $q(\theta)$ is *any* density on the model parameters (Neal and Hinton, 1998). Equation B.1 rests on Jensen's inequality that follows from the concavity of the log function, which renders the log of an integral greater than the integral of the log. F corresponds to the [negative] free energy in statistical thermodynamics and comprises two terms; the energy and entropy. The **EM** algorithm alternates between maximising F, and implicitly the likelihood of the data, with respect to the distribution $q(\theta)$ and the hyperparameters λ, holding the other fixed

$$\textbf{E}\text{-step: } q(\theta) \leftarrow \max_{q} F(q(\theta), \lambda)$$

$$\mathbf{M}\text{-step: } \lambda \leftarrow \max_{\lambda} F(q(\theta), \lambda)$$

This iterative alternation performs a co-ordinate ascent on *F*. It is easy to show that the maximum in the **E**-step obtains when $q(\theta) = p(\theta \mid y, \lambda)$, at which point Eq. B.1 becomes an equality. The **M**-step finds the maximum likelihood (ML) estimate of the hyperparameters, *i.e.* the values of λ that maximise $p(y \mid \lambda)$ by integrating $\ln p(\theta, y \mid \lambda) = \ln p(y \mid \theta, \lambda) + \ln p(\theta \mid \lambda)$ over the parameters, using the current estimate of their conditional distribution. In short, the **E**-step computes sufficient statistics (in our case the conditional mean and covariance) of the unobserved parameters to enable the **M**-step to optimise the hyperparameters, in a maximum likelihood sense. These new hyperparameters re-enter into the estimation of the conditional density and so on until convergence.

The E-Step

For linear models, under Gaussian (*i.e.* parametric) assumptions, the **E**-step corresponds to evaluating the conditional mean and covariance

$$y = X\theta + \varepsilon$$

$$\overline{y} = \begin{bmatrix} y - X\theta \\ \eta_\theta \end{bmatrix} \quad \overline{X} = \begin{bmatrix} X \\ I \end{bmatrix} \quad \overline{C}_\varepsilon = \begin{bmatrix} \sum \lambda_i Q_i & 0 \\ 0 & C_\theta \end{bmatrix}. \tag{B.2}$$

$$\eta_{\theta|y} = C_{\theta|y} \overline{X}^T \overline{C}_\varepsilon^{-1} \overline{y}$$
$$C_{\theta|y} = (\overline{X}^T \overline{C}_\varepsilon^{-1} \overline{X})^{-1}$$

Where the prior and conditional densities are $p(\theta) = N(\eta_\theta, C_\theta)$ and $q(\theta) = N(\eta_{\theta|y}, C_{\theta|y})$. This compact form is a result of absorbing the priors into the errors by augmenting the linear system. The same augmentation is used to reduce hierarchal models, with empirical priors to their non-hierarchical form. Under local

linearity assumptions, non-linear models can be reduced to a linear form. The resulting conditional density is used to estimate the hyperparameters of the covariance components in the **M**-step:

The M-Step

Given that we can reduce the problem to estimating the error covariances of the augmented system in B.2; we need to estimate the hyperparameters of the error covariances (which contain the prior covariances). Specifically, we require the hyperparameters that maximise the first term of the free energy (i.e., the energy) because the entropy does not depend on the hyperparameters. For linear systems the free energy is given by (ignoring constants)

$$\begin{aligned}\log p(\theta, y|\lambda) &= -\tfrac{1}{2}\ln|C_\varepsilon| - \tfrac{1}{2}(\bar{y}-\bar{X}\theta)^T C_\varepsilon^{-1}(\bar{y}-\bar{X}\theta).\\ \int q(\theta)\ln p(\theta, y|\lambda)d\theta &= -\tfrac{1}{2}\ln|C_\varepsilon| - \tfrac{1}{2}r^T C_\varepsilon^{-1} r - \tfrac{1}{2}tr\{C_{\theta|y}\bar{X}^T C_\varepsilon^{-1}\bar{X}\}\\ \int q(\theta)\log q(\theta) &= -\tfrac{1}{2}\ln|C_{\theta|y}|\end{aligned} \tag{B.3}$$

$$F = \tfrac{1}{2}\ln|C_\varepsilon^{-1}| - \tfrac{1}{2}r^T C_\varepsilon^{-1} r - \tfrac{1}{2}tr\{C_{\theta|y}\bar{X}^T C_\varepsilon^{-1}\bar{X}\} + \tfrac{1}{2}\ln|C_{\theta|y}|$$

where the residuals $r = \bar{y} - \bar{X}\eta_{\theta|y}$. By taking derivatives with respect to the error covariance we get

$$\frac{\partial F}{\partial C_\varepsilon^{-1}} = \tfrac{1}{2}C_\varepsilon - \tfrac{1}{2}rr^T - \tfrac{1}{2}\bar{X}C_{\theta|y}\bar{X}^T. \tag{B.4}$$

When the hyperparameters maximise the free energy this gradient is zero and

$$C(\lambda)_\varepsilon = rr^T + \bar{X}C_{\theta|y}\bar{X}^T \tag{B.5}$$

(*c.f.* (Dempster, 1981) p350). This means that the ReML error covariance estimate has two components: that due to differences between the data and its conditional

prediction and another due to the variation of the parameters about their conditional mean; *i.e.*, their conditional uncertainty. This is not a closed form expression for the unknown covariance because the conditional covariance is a function of the hyperparameters. To find the ReML hyperparameters one usually adopts a Fisher scoring scheme, using the first and expected second partial derivatives of the free energy.

$$
\begin{aligned}
\Delta\lambda &= -E\left(\frac{\partial^2 F}{\partial\lambda_{ij}^2}\right)^{-1}\frac{\partial F}{\partial\lambda_i} \\
\frac{\partial F}{\partial\lambda_i} &= tr\left(\frac{\partial F}{\partial C_\varepsilon^{-1}}C_\varepsilon^{-1}Q_iC_\varepsilon^{-1}\right) = -\tfrac{1}{2}tr\{PQ_i\} + \tfrac{1}{2}\overline{y}^T P^T Q_i P\overline{y} \\
\frac{\partial^2 F}{\partial\lambda_{ij}^2} &= \tfrac{1}{2}tr\{PQ_iPQ_j\} - \overline{y}^T PQ_iPQ_jP\overline{y} \qquad . \qquad \text{(B.6)} \\
E\left(\frac{\partial^2 F}{\partial\lambda_{ij}^2}\right) &= -\tfrac{1}{2}tr\{PQ_iPQ_j\}
\end{aligned}
$$

$$
P = C_\varepsilon^{-1} - C_\varepsilon^{-1}\overline{X}C_{\theta|y}\overline{X}^T C_\varepsilon^{-1}
$$

Fisher scoring corresponds to augmenting a Gauss-Newton scheme by replacing the second derivative or curvature with its expectation. The curvature or Hessian is referred to as Fisher's Information matrix[12] and encodes the conditional prediction of the hyperparameters. In this sense, the Information matrix has a close connection to the degrees of freedom in classical statistics. The gradient can be computed efficiently by capitalising on any sparsity structure in the constraints and by bracketing the multiplications appropriately. This scheme is general in that it accommodates almost any form for the covariance through a Taylor expansion of $C\{\lambda\}_\varepsilon$.

[12] The derivation of the expression for the Information matrix uses standard results from linear algebra and is most easily seen by differentiating the gradient, noting

$$
\frac{\partial P}{\partial\lambda_j} = -PQ_jP
$$

and taking the expectation, using

$$
E(tr(PQ_iP\overline{yy}^T PQ_j)) = tr\{PQ_iPC_\varepsilon PQ_j\} = tr\{PQ_iPQ_j\}
$$

Once the hyperparameters have been updated they enter the **E**-step as a new error covariance estimate to give new conditional moments, which, in turn enter the **M**-step and so on until convergence.

It should be noted that the search for the maximum of F does not have to employ Fisher scoring or indeed the parameterisation of C_ε used above. Other search procedures such as quasi-Newton searches are commonly employed (Fahrmeir, 1994). (Harville, 1977) originally considered Newton-Raphson and scoring algorithms, and (Laird and Ware, 1982) recommend several versions of **EM**. One limitation of the linear hyper-parameterisation described above is that does not guarantee that C_ε is positive definite. This is because the hyperparameters can take negative values with extreme degrees of non-sphericity. The **EM** algorithm employed by *multistat* (Worsley et al., 2002), for variance component estimation in multi-subject fMRI studies, uses a slower but more stable algorithm that ensures positive definite covariance estimates.

Appendix C

Approximations to the log model evidence

With the exception of some special cases (e.g., linear models), the integral expression for the model evidence (Eq. C.1) is analytically intractable and numerically difficult to compute. Under these circumstances, people generally adopt a bound approach where, instead of evaluating the integral above, one optimises a bound on the integral using iterative sampling or analytic techniques. The most common approach of the latter kind is variational Bayes. In this framework, one posits an approximating conditional or posterior density on the unknown parameters, *q(θ)*, and optimises this density with respect to a free-energy bound, *F*, on the log-evidence:

$$F = \log p\left(y|m\right) - KL\left[q(\theta), p\left(\theta|y,m\right)\right] \tag{C.1}$$

Because of its relation to variational calculus and Gibb's free-energy in statistical physics, this free-energy bound *F* is often referred to as the "negative free-energy" or "variational free-energy" (Friston et al., 2007; MacKay, 2003; Neal and Hinton, 1998). Its second term is the Kullback–Leibler (KL) divergence (Kullback and Leibler, 1951) between the approximating posterior density *q(θ)* and the true posterior $p(\theta \mid y, m)$, which is always positive (or zero when *q(θ)* becomes identical to $p(\theta \mid y, m)$). By iterative optimisation, the negative free-energy *F* is made an increasingly tighter lower bound on the desired log-evidence, $\ln p(y|m)$; as a consequence, the KL divergence between the approximating and true posterior is minimised. There are a number of approximations that are used when specifying the form of *q(θ)*. These include the ubiquitous mean-field approximation, where various sets of unknown parameters are assumed to be independent, so that the conditional density can be factorised. A common example here would be a bipartition into the regression coefficients of a general linear model and the parameters controlling random effects or error variance. Another common approximation within the mean-field framework is to assume that the conditional density is multivariate Gaussian. This is also known as the Laplace approximation, a full treatment of which can be found in (Friston et al., 2007).

Negative free energy is a lower bound on the log model evidence

The relation of the negative free energy F to the log model evidence, $\log p(y|m)$, can be derived by using an (arbitrary) approximating posterior density $q(\theta)$ to decompose $p(y|m)$ into two components, *i.e.*, F and the Kullback-Leibler divergence (KL) between the true posterior $p(y|\, m)$ and the approximating posterior $q(\theta)$:

$$\begin{aligned}\log p(y|m) &= \int q(\theta)\log p(y|m)d\theta \\ &= \int q(\theta)\log\frac{p(y,\theta|m)q(\theta)}{p(\theta|y,m)q(\theta)}d\theta \\ &= \int q(\theta)\log\frac{p(y,\theta|m)}{q(\theta)}d\theta + \int q(\theta)\log\frac{q(\theta)}{p(\theta|y,m)}d\theta \\ &= F + KL\left[q(\theta, p(\theta|y,m))\right]\end{aligned} \quad . \tag{C.2}$$

The KL divergence is an asymmetric measure of the differences between two probability densities (Kullback and Leibler, 1951). If the approximating posterior matches the true posterior density precisely, then $KL[q(\theta), p(\theta|y,m)] = 0$. This demonstrates that the negative free energy F is a lower bound on the log-evidence and can therefore be used as a criterion for model comparison. This makes the assumption that the KL divergence term is not drastically different across models (*i.e.*, the tightness of the bound is similar under different models). For models like ours, with informed priors that lead to well-behaved posterior densities, this assumption is unlikely to be a strong one.

Appendix D

Stability analysis of neuronal networks for simple equations

The solution to a first order differential equation $\frac{d\vec{X}}{dt} = \vec{A}\vec{X} + \vec{B}$ is obtained by solving for the equilibrium state $\vec{X}_{eq} = -\vec{A}^{-1}\vec{B}$. One can then determine the two eigenvalues, λ_1 and λ_2, of the characteristic equation $\left|\vec{A} - \lambda\vec{I}\right| = 0$, where I is the identity matrix. Assuming, $\lambda_1 \neq \lambda_2$ the solution is $\vec{X} = \begin{pmatrix} a_1 e^{\lambda_1 t} \\ b_1 e^{\lambda_2 t} \end{pmatrix} + \begin{pmatrix} a_2 e^{\lambda_2 t} \\ b_2 e^{\lambda_2 t} \end{pmatrix} - \vec{X}_{eq}$, where a_1, a_2, b_1, b_2 are constants.

If all eigenvalues of the linear system have negative real parts, the system is asymptotically stable. If the linear system has at least one eigenvalue with a positive real part, the system is unstable. In addition, the type of equilibrium point for the system, can be spiral point, node or a saddle point. The same holds [locally] for nonlinear systems.

If we consider a two node network, written in matrix form as $A = \begin{bmatrix} a & c \\ b & a \end{bmatrix}$. We have the associated stability analysis represented in the figure below:

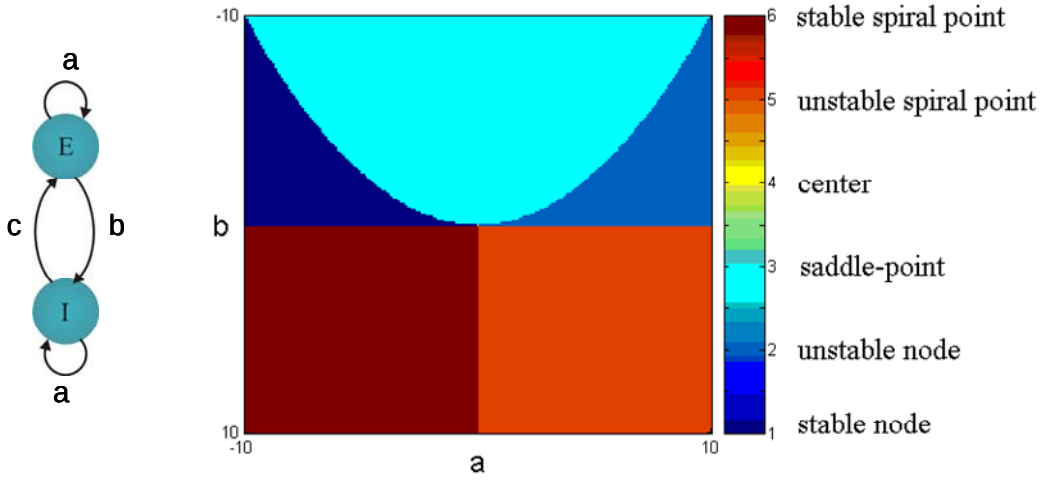


Figure D.1: 2D stability diagram for different equilibrium points for a two node network.

Appendix E

Contribution of variance over states and thresholds

The predicted firing rate of a population is the expectation of the step or Heaviside function of depolarisation, over both the states and the threshold probability density functions (Equation 4.2 in the main text):

$$\left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_{p(w)p(x_1)} = \iint H(x_1 - w)\, p(x_1, w) dx_1 dw \,. \tag{E.1}$$

Assuming x_1 and w are independent and normally distributed; $z = x_1 - w$ has a Gaussian distribution; $p(z) = N(\mu_z, \Sigma_z) = N(\mu_x - \mu_w, \Sigma_x + \Sigma_w)$ and Equation A.1 can be written as:

$$\begin{aligned} \left\langle H(x_1^{(j)} - w^{(j)}) \right\rangle_{p(w)p(x_1)} &= \int H(z) dz \\ &= \int_{z \geq 0} p(z) dz \\ &\approx \frac{1}{1 + \exp(\rho(\Sigma_z)(\mu_x - \mu_w))} \end{aligned} \,. \tag{E.2}$$

This means the expected firing rate remains a function of the sufficient statistics of the population and retains the same form as Equation 4.3. Furthermore, it shows that for any given value of the slope parameter, $\rho(\Sigma_z)$ the implicit variance

$$\rho(\Sigma_z)^{-1} = \Sigma_z(\rho) = \Sigma_x + \Sigma_w \geq \Sigma_x \,. \tag{E.3}$$

is always greater than the population variance on neuronal states. This means we always overestimate the proportion of supra-threshold neurons that contribute to the firing because $\Sigma_z(\rho)$ is an overestimate of the population variance. In other words, the 12% estimate from Figure 4.7 is an upper bound on the actual proportion of firing neurons.

Appendix F

Neural-field models

Neural-mass models can be generalised to neural-field models by making the modes a function of space, thereby furnishing wave-equations that describe the spatiotemporal evolution of neuronal states over the cortical surface. An important extension of neural-mass models speaks to the fact that neuronal dynamics play out on a spatially extended cortical sheet. In other words, states like the depolarisation of an excitatory ensemble in the granular layer of cortex can be regarded as a *continuum* or *field*, which is a function of space r and time $\mu(t) \to \mu(r,t)$. This allows one to formulate the dynamics of the expected field in terms of partial differential equations in space and time. These are essentially wave equations that accommodate, gracefully, lateral interactions, which are generally assumed to be stationary across the cortical sheet. Neural-field models were among the first mean-field models of neuronal dynamics (Wilson and Cowan, 1972). Key forms for neural-field equations were proposed and analysed by (Nunez, 1974) and (Amari, 1975, 1977). These models were generalised by (Jirsa and Haken, 1997) who, critically, considered delays in the propagation of spikes over space. The introduction of propagation delays, leads to dynamics that are very reminiscent of those observed empirically. Typically, neural-field models can be construed as a spatiotemporal convolution (*c.f.*, Eq. 5.6) that can be written in terms of a Green function

$$\begin{aligned} \mu_V(r,t) &= \int G(r-r',t-t')\varsigma(\mu_V(r,t))dt'dr' \\ G(r-r',t-t') &= \delta\left(t-t'-\tfrac{1}{c}\,|\,r-r'\,|\right)\exp(-\tfrac{1}{\gamma}\,|\,r-r'\,|) \end{aligned}. \qquad \text{(F.1)}$$

Where $|\,r-r'\,|$ is the distance between r and r', c is the speed of spike propagation and γ controls the spatial decay of lateral interactions. The corresponding second-order equations of motion are a neural wave equation (see Daunizeau et al., 2009)

$$\left(\frac{\partial^2}{\partial t^2} + 2\kappa\frac{\partial}{\partial t} + \kappa^2 - \frac{3}{2}c^2\nabla^2\right)\mu_V = c\kappa\varsigma(\mu_V). \qquad \text{(F.2)}$$

Where $\kappa = c/\gamma$. The formal similarity with the neural-mass model in Eq. 5.15 is self-evident. These sorts of models have been extremely useful in modelling spatiotemporally extended dynamics (*e.g.*, (Breakspear et al., 2003a; Liley and Bojak, 2005)).

BIBLIOGRAPHY

Absher, J.R., Benson, D. F. , 1993. Disconnection syndromes: an overview of Geschwind's contributions. Neurology 43, 862-867.

Acs, F., Greenlee, M.W., 2008. Connectivity modulation of early visual processing areas during covert and overt tracking tasks. NeuroImage 41, 380-388.

Aertsen, A., Preissl, H., 1991. Dynamics of activity and connectivity in physiological neuronal Networks. Nonlinear Dynamics and Neuronal Networks 2, 281-301.

Allen, P., Mechelli, A., Stephan, K.E., Day, F., Dalton, J., Williams, S., McGuire, P.K., 2008. Fronto-temporal interactions during overt verbal initiation and suppression. J Cogn Neurosci 20, 1656-1669.

Allison, T., McCarthy, G., Wood, C.C., Jones, S.J., 1991. Potentials evoked in human and monkey cerebral cortex by stimulation of the median nerve. A review of scalp and intracranial recordings. Brain 114 (Pt 6), 2465-2503.

Amari, S., 1972. Characteristics of random nets of analog neuron-like elements. Ieee Transactions on Systems Man and Cybernetics SMC2, 643-&.

Amari, S., 1975. Homogeneous nets of neuron-like elements. Biol Cybern 17, 211-220.

Amari, S., 1977. Dynamics of pattern formation in lateral-inhibition type neural fields. Biol Cybern 27, 77-87.

Ashburner, J., Friston, K.J., 2000. Voxel-based morphometry--the methods. NeuroImage 11, 805-821.

Baillet, S., Mosher, J.C., Leahy, R.M., 2001. Electromagnetic brain mapping. Ieee Signal Processing Magazine 18, 14-30.

Baumgartner, R., Scarth, G., Teichtmeister, C., Somorjai, R., Moser, E., 1997. Fuzzy clustering of gradient-echo functional MRI in the human visual cortex. Part I: reproducibility. Journal of Magnetic Resonance Imaging 7.

Berry, D., Hochberg, Y., 1999. Bayesian perspectives on multiple comparisons. Journal of Statistical Planning and Inference 82, 215-227.

Bertrand, O., Tallon-Baudry, C., 2000. Oscillatory gamma activity in humans: a possible role for object representation. Int J Psychophysiol 38, 211-223.

Beurle, R.L., 1956. Properties of a mass of cells capable of regenerating pulses. Philosophical Transactions of the Royal Society of London Series B-Biological Sciences 240, 55-87.

Biswal, B., Zerrin Yetkin, F., Haughton, V., Hyde, J., 1995. Functional connectivity in the motor cortex of resting human brain using echo-planar MRI. Magnetic Resonance in Medicine 34.

Bitan, T., Booth, J.R., Choy, J., Burman, D.D., Gitelman, D.R., Mesulam, M.M., 2005. Shifts of effective connectivity within a language network during rhyming and spelling. J Neurosci 25, 5397-5403.

Bliss, T.V., Lomo, T., 1973. Long-lasting potentiation of synaptic transmission in the dentate area of the anaesthetized rabbit following stimulation of the perforant path. J Physiol 232, 331-356.

Boynton, G., Engel, S., Glover, G., Heeger, D., 1996. Linear systems analysis of functional magnetic resonance imaging in human V1. Journal of Neuroscience 16, 4207-4221.

Brandes, U., Erlebach, T, 2005. Network Analysis. Springer, Berlin.

Breakspear, M., Roberts, J.A., Terry, J.R., Rodrigues, S., Mahant, N., Robinson, P.A., 2006. A unifying explanation of primary generalized seizures through nonlinear brain modelling and bifurcation analysis. Cereb Cortex 16, 1296-1313.

Breakspear, M., Terry, J., Friston, K., 2003a. Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network: Computation in Neural Systems 14, 703-732.

Breakspear, M., Terry, J.R., Friston, K.J., 2003b. Modulation of excitatory synaptic coupling facilitates synchronization and complex dynamics in a biophysical model of neuronal dynamics. Network 14, 703-732.

Brodmann, 1909. Vergleichende Lokalisationslehre der Grosshirnrinde in ihren Prinzipien dargestellt auf Grund des Zellenbaues. J.A. Barth, Leipzig.

Büchel, C., Friston, K.J., 1997. Modulation of connectivity in visual pathways by attention: cortical interactions evaluated with structural equation modelling and fMRI. Cereb Cortex 7, 768-778.

Büchel, C., Friston, K.J., 1998. Dynamic changes in effective connectivity characterized by variable parameter regression and Kalman filtering. Hum Brain Mapp 6, 403-408.

Büchel, C., Wise, R., Mummery, C., Poline, J., Friston, K., 1996. Nonlinear regression in parametric activation studies. Neuroimage 4, 60.

Buchner, H., Adams, L., Muller, A., Ludwig, I., Knepper, A., Thron, A., Niemann, K., Scherg, M., 1995. Somatotopy of human hand somatosensory cortex revealed by dipole source analysis of early somatosensory evoked potentials and 3D-NMR tomography. Electroencephalogr Clin Neurophysiol 96, 121-134.

Buxton, R.B., Wong, E.C., Frank, L.R., 1998. Dynamics of blood flow and oxygenation changes during brain activation: the balloon model. Magn Reson Med 39, 855-864.

Cajal, S.R., 1909. Histologie du systéme nerveux de l'homme et des vertebras. Maloine, Paris.

Canolty, R.T., Edwards, E., Dalal, S.S., Soltani, M., Nagarajan, S.S., Kirsch, H.E., Berger, M.S., Barbaro, N.M., Knight, R.T., 2006. High gamma power is phase-locked to theta oscillations in human neocortex. Science 313, 1626-1628.

Casti, A.R., Omurtag, A., Sornborger, A., Kaplan, E., Knight, B., Victor, J., Sirovich, L., 2002. A population study of integrate-and-fire-or-burst neurons. Neural Comput 14, 957-986.

Chen, C.C., Kiebel, S.J., Friston, K.J., 2008. Dynamic causal modelling of induced responses. NeuroImage 41, 1293-1312.

Chizhov, A.V., Graham, L.J., 2007. Population model of hippocampal pyramidal neurons, linking a refractory density approach to conductance-based neurons. Phys Rev E Stat Nonlin Soft Matter Phys 75, 011924.

Chizhov, A.V., Graham, L.J., Turbin, A.A., 2006. Simulation of neural population dynamics with a refractory density approach and a conductance-based threshold neuron model. Neurocomputing 70, 252-262.

Chrobak, J.J., Buzsaki, G., 1998. Gamma oscillations in the entorhinal cortex of the freely behaving rat. J Neurosci 18, 388-398.

Chumbley, J.R., Friston, K.J., Fearn, T., Kiebel, S.J., 2007. A Metropolis-Hastings algorithm for dynamic causal models. NeuroImage 38, 478-487.

Connor JA and Stevens CF. "Prediction of repetitive firing behaviour from voltage clamp data on an isolated neurone soma." J Physiol. 1971 Feb;213(1):31-53.

Cooke, S.F., Bliss, T.V., 2006. Plasticity in the human central nervous system. Brain 129, 1659-1673.

Cudeck, R., 2002. Structural Equation Modelling: Present and Future. Lincolnwood, IL.

Daunizeau, J., Friston, K.J., Kiebel, S.J., 2009. Variational Bayesian identification and prediction of stochastic nonlinear dynamic causal models. Physica D: nonlinear phenomena *in revision*.

Daunizeau, J., Kiebel, S.J., Friston, K.J., 2009. Dynamic causal modelling of distributed responses. Neuroimage doi:10.1016/j.physletb.2003.10.071.

David, O., Friston, K.J., 2003. A neural mass model for MEG/EEG: coupling and neuronal dynamics. NeuroImage 20, 1743-1755.

David, O., Harrison, L., Friston, K.J., 2005. Modelling event-related responses in the brain. NeuroImage 25, 756-770.

David, O., Kiebel, S.J., Harrison, L.M., Mattout, J., Kilner, J.M., Friston, K.J., 2006a. Dynamic causal modelling of evoked responses in EEG and MEG. NeuroImage 30, 1255-1272.

David, O., Kilner, J.M., Friston, K.J., 2006b. Mechanisms of evoked and induced responses in MEG/EEG. NeuroImage 31, 1580-1591.

Dayan, P., Abbott, L.F., 2001. Theoretical Neuroscience: Computational and Mathematical Modelling of Neural Systems. The MIT Press.

Deco, G., Marti, D., 2007. Extended method of moments for deterministic analysis of stochastic multistable neurodynamical systems. Phys Rev E Stat Nonlin Soft Matter Phys 75, 031913.

Deco, G., Jirsa, V.K., Robinson, P.A., Breakspear, M., Friston, K., 2008. The dynamic brain: from spiking neurons to neural masses and cortical fields. PLoS Comput Biol 4, e1000092.

Dempster, A.P., Rubin, D. B., Tsutakawa, R., K. , 1981. Estimation in covariance component models. Journal of the American Statistical Association 76, 341-353.

den Ouden, H.E., Friston, K.J., Daw, N.D., McIntosh, A.R., Stephan, K.E., 2008. A Dual Role for Prediction Error in Associative Learning. Cereb Cortex.

Destexhe, A., Pare, D., 1999. Impact of network activity on the integrative properties of neocortical pyramidal neurons in vivo. J Neurophysiol 81, 1531-1547.

Doeller, C.F., Opitz, B., Mecklinger, A., Krick, C., Reith, W., Schroger, E., 2003. Prefrontal cortex involvement in preattentive auditory deviance detection: neuroimaging and electrophysiological evidence. NeuroImage 20, 1270-1282.

Doiron, B., Rinzel, J., Reyes, A., 2006. Stochastic synchronization in finite size spiking networks. Phys Rev E Stat Nonlin Soft Matter Phys 74, 030903.

Draganski, B., Gaser, C., Kempermann, G., Kuhn, H.G., Winkler, J., Büchel, C., May, A., 2006. Temporal and spatial dynamics of brain structure changes during extensive learning. J Neurosci 26, 6314-6317.

Eggert, J., van Hemmen, J.L., 2001. Modelling neuronal assemblies: theory and implementation. Neural Comput 13, 1923-1974.

Eickhoff, S.B., Dafotakis, M., Grefkes, C., Shah, N.J., Zilles, K., Piza-Katzer, H., 2008. Central adaptation following heterotopic hand replantation probed by fMRI and effective connectivity analysis. Exp Neurol 212, 132-144.

Elbert, T., Ray, W.J., Kowalik, Z.J., Skinner, J.E., Graf, K.E., Birbaumer, N., 1994. Chaos and physiology: deterministic chaos in excitable cell assemblies. Physiol Rev 74, 1-47.

Escera, C., Yago, E., Corral, M.J., Corbera, S., Nunez, M.I., 2003. Attention capture by auditory significant stimuli: semantic analysis follows attention switching. Eur J Neurosci 18, 2408-2412.

Ethofer, T., Anders, S., Erb, M., Herbert, C., Wiethoff, S., Kissler, J., Grodd, W., Wildgruber, D., 2006. Cerebral pathways in processing of affective prosody: a dynamic causal modelling study. NeuroImage 30, 580-587.

Fahrmeir, L., Tutz, G. , 1994. Multivariate statistical modelling based on generalised linear models. Springer, New York.

Fairhall, S.L., Ishai, A., 2007. Effective connectivity within the distributed cortical network for face perception. Cereb Cortex 17, 2400-2406.

Felleman, D.J., Van Essen, D.C., 1991. Distributed hierarchical processing in the primate cerebral cortex. Cereb Cortex 1, 1-47.

Foster, B.L., Bojak, I., Liley, D.T., 2008. Population based models of cortical drug response: insights from anaesthesia. Cogn Neurodyn 2, 283-296.

Frank, T.D., 2004. Nonlinear Fokker-Planck Equations: Fundamentals and Applications. Springer, Berlin.

Frank, T.D., Daffertshofer, A., Beek, P.J., 2001. Multivariate Ornstein-Uhlenbeck processes with mean-field dependent coefficients: application to postural sway. Phys Rev E Stat Nonlin Soft Matter Phys 63, 011905.

Freeman, W.J., 1975. Mass Action in the Nervous System. New York: Academic Press.

Freeman, W.J., 1978. Models of the dynamics of neural populations. Electroencephalogr Clin Neurophysiol Suppl, 9-18.

Fricker, D., Verheugen, J.A., Miles, R., 1999. Cell-attached measurements of the firing threshold of rat hippocampal neurones. J Physiol 517 (Pt 3), 791-804.

Friston, K., 2002a. Beyond phrenology: what can neuroimaging tell us about distributed circuitry? Annu Rev Neurosci 25, 221-250.

Friston, K., Chu, C., Mourão-Miranda, J., Hulme, O., Rees, G., Penny, W., Ashburner, J., 2008. Bayesian decoding of brain images. NeuroImage 39, 181-205.

Friston, K., Frith, C., Fletcher, P., Liddle, P., Frackowiak, R., 1996a. Functional topography: multidimensional scaling and functional connectivity in the brain. Cerebral Cortex 6, 156-164.

Friston, K., Frith, C., Liddle, P., Frackowiak, R., 1991. Comparing functional(PET) images: the assessment of significant change. Journal of Cerebral Blood Flow and Metabolism 11, 690-699.

Friston, K., Mattout, J., Trujillo-Barreto, N., Ashburner, J., Penny, W., 2007. Variational free energy and the Laplace approximation. NeuroImage 34, 220-234.

Friston, K., Poline, J., Holmes, A., Frith, C., Frackowiak, R., 1996b. A multivariate analysis of PET activation studies. Human Brain Mapping 4.

Friston, K.J., 1994. Functional and effective connectivity in neuroimaging: a synthesis. Hum Brain Mapp 2, 56-78.

Friston, K.J., 1997. Another neural code? NeuroImage 5, 213-220.

Friston, K.J., 2002b. Bayesian estimation of dynamical systems: an application to fMRI. NeuroImage 16, 513-530.

Friston, K.J., 2003. Volterra kernels and connectivity. Human Brain Function. Academic Press.

Friston, K.J., 2005. A theory of cortical responses. Philos Trans R Soc Lond B Biol Sci 360, 815-836.

Friston, K.J., Buchel, C., 2000. Attentional modulation of effective connectivity from V2 to V5/MT in humans. Proc Natl Acad Sci U S A 97, 7591-7596.

Friston, K.J., Buechel, C., Fink, G.R., Morris, J., Rolls, E., Dolan, R.J., 1997. Psychophysiological and modulatory interactions in neuroimaging. NeuroImage 6, 218-229.

Friston, K.J., Frith C. D., Frackowiak R. S. J. , 1993. Time-dependent changes in effective connectivity measured with PET. Human Brain Mapp 1, 69-79.

Friston, K.J., Frith, C.D., Liddle, P.F., Frackowiak, R.S., 1993. Functional connectivity: the principal-component analysis of large (PET) data sets. J Cereb Blood Flow Metab 13, 5-14.

Friston, K.J., Harrison, L., Penny, W., 2003. Dynamic causal modelling. NeuroImage 19, 1273-1302.

Friston, K.J., Holmes A.P., Worsley K.J., Poline J.B., Frith C.D., Frackowiak, R.S.J., 1995a. Statistical Parametric Maps in functional imaging: A general linear approach. Human Brain Mapp 2, 189-210.

Friston, K.J., Mechelli, A., Turner, R., Price, C.J., 2000. Nonlinear responses in fMRI: the Balloon model, Volterra kernels, and other hemodynamics. NeuroImage 12, 466-477.

Friston, K.J., Penny, W., Phillips, C., Kiebel, S., Hinton, G., Ashburner, J., 2002. Classical and Bayesian inference in neuroimaging: theory. NeuroImage 16, 465-483.

Friston, K.J., Ungerleider L. G., Jezzard P. , 1995b. Charcterizing modulatory interactions between V1 and V2 in human cortex: a new treatment of functional MRI data. Hum Brain Mapp 2, 211-224.

Fukai, T., 1999. Sequence generation in arbitrary temporal patterns from theta-nested gamma oscillations: a model of the basal ganglia-thalamo-cortical loops. Neural Netw 12, 975-987.

Galan, R.F., Ermentrout, G.B., Urban, N.N., 2007. Stochastic dynamics of uncoupled neural oscillators: Fokker-Planck studies with the finite element method. Phys Rev E Stat Nonlin Soft Matter Phys 76, 056110.

Garrido, M.I., Friston, K.J., Kiebel, S.J., Stephan, K.E., Baldeweg, T., Kilner, J.M., 2008. The functional anatomy of the MMN: a DCM study of the roving paradigm. NeuroImage 42, 936-944.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Friston, K.J., 2007a. Evoked brain responses are generated by feedback loops. Proc Natl Acad Sci U S A 104, 20961-20966.

Garrido, M.I., Kilner, J.M., Kiebel, S.J., Stephan, K.E., Friston, K.J., 2007b. Dynamic causal modelling of evoked potentials: a reproducibility study. NeuroImage 36, 571-580.

Genovese, C.R., Lazar, N.A., Nichols, T., 2002. Thresholding of statistical maps in functional neuroimaging using the false discovery rate. NeuroImage 15, 870-878.

Gerstner, W., 2001. A framework for spiking neuron models: The spike response model. Springer Berlin / Heidelberg.

Gerstner, W., Kistler, W. M., 2002. Spiking Neuron Models Single Neurons, Populations, Plasticity. Cambridge University Press.

Giard, M.H., Perrin, F., Pernier, J., Bouchet, P., 1990. Brain generators implicated in the processing of auditory stimulus deviance: a topographic event-related potential study. Psychophysiology 27, 627-640.

Goebel, R., Roebroeck, A., Kim, D.S., Formisano, E., 2003. Investigating directed cortical interactions in time-resolved fMRI data using vector autoregressive modelling and Granger causality mapping. Magn Reson Imaging 21, 1251-1261.

Granger, C., 1969. Investigating causal relations by econometric models and cross-spectral methods. Econometrica: Journal of the Econometric Society, 424-438.

Griffith, J.S., 1963. A field theory of neural nets: I. Derivation of field equations. Bull Math Biophys 25, 111-120.

Griffith, J.S., 1965. A field theory of neural nets. II. Properties of the field equations. Bull Math Biophys 27, 187-195.

Griffiths, T.D., Kumar, S., Warren, J.D., Stewart, L., Stephan, K.E., Friston, K.J., 2007. Approaches to the cortical analysis of auditory objects. Hear Res 229, 46-53.

Grol, M.J., Majdandzic, J., Stephan, K.E., Verhagen, L., Dijkerman, H.C., Bekkering, H., Verstraten, F.A., Toni, I., 2007. Parieto-frontal connectivity during visually guided grasping. J Neurosci 27, 11877-11887.

Haenschel, C., Uhlhaas, P.J., Singer, W., 2007. Synchronous oscillatory activity and working memory in schizophrenia. Pharmacopsychiatry 40, S54-S61.

Harrison, L., Penny, W.D., Friston, K., 2003. Multivariate autoregressive modelling of fMRI time series. NeuroImage 19, 1477-1491.

Harrison, L.M., David, O., Friston, K.J., 2005. Stochastic models of neuronal dynamics. Philos Trans R Soc Lond B Biol Sci 360, 1075-1091.

Harville, D.A., 1977. Maximim Likelihood Approaches to Variance Component and to Related Problems. Journal of the American Statistical Association 72, 320-338.

Hasegawa, H., 2003a. Dynamical mean-field theory of noisy spiking neuron ensembles: Application to the Hodgkin-Huxley model. Physical review. E, Statistical, nonlinear, and soft matter physics 68, 41909-41909.

Hasegawa, H., 2003b. Dynamical mean-field theory of spiking neuron ensembles: Response to a single spike with independent noises. Phys. Rev. E 67, 041903.041901-041903.041919.

Hasegawa, H., 2004. Dynamical mean-field approximation to small-world networks of spiking neurons: from local to global and/or from regular to random couplings. Phys Rev E Stat Nonlin Soft Matter Phys 70, 066107.

Hasegawa, H., 2006. N-dependent multiplicative-noise contributions in finite N-unit Langevin models: Augmented moment approach. Journal of the Physical Society of Japan 75.

Hasegawa, H., 2007. Stationary and dynamical properties of finite N-unit Langevin models subjected to multiplicative noises. Physica a-Statistical Mechanics and Its Applications 374, 585-599.

Haskell, E., Nykamp, D.Q., Tranchina, D., 2001. Population density methods for large-scale modelling of neuronal networks with realistic synaptic kinetics: cutting the dimension down to size. Network 12, 141-174.

Heim, S., Eickhoff, S.B., Ischebeck, A.K., Friederici, A.D., Stephan, K.E., Amunts, K., 2009. Effective connectivity of the left BA 44, BA 45, and inferior temporal gyrus during lexical and phonological decisions identified with DCM. Hum Brain Mapp 30, 392-402.

Hocking, A.B., Levy, W. B., 2007. Theta-modulated input reduces intrinsic gamma oscillation in a hippocampal model. Neurocomputing 70, 2074-2078.

Hodgkin, A.L., Huxley, A.F., 1952. A quantitative description of membrane current and its application to conduction and excitation in nerve. J Physiol 117, 500-544.

Horwitz, B., Tagamets, M.A., McIntosh, A.R., 1999. Neural modelling, functional brain imaging, and cognition. Trends Cogn Sci 3, 91-98.

Izhikevich, E., 2007. Dynamical Systems in Neuroscience: The Geometry of Excitability and Bursting. MIT Press.

Jääskeläinen, I.P., Ahveninen, J., Bonmassar, G., Dale, A.M., Ilmoniemi, R.J., Levanen, S., Lin, F.H., May, P., Melcher, J., Stufflebeam, S., Tiitinen, H., Belliveau,

J.W., 2004. Human posterior auditory cortex gates novel sounds to consciousness. Proc Natl Acad Sci U S A 101, 6809-6814.

Jansen, B.H., Kavaipatti, A.B., Markusson, O., 2001. Evoked potential enhancement using a neurophysiologically-based model. Methods Inf Med 40, 338-345.

Jansen, B.H., Rit, V.G., 1995. Electroencephalogram and visual evoked potential generation in a mathematical model of coupled cortical columns. Biol Cybern 73, 357-366.

Jemel, B., Achenbach, C., Muller, B.W., Ropcke, B., Oades, R.D., 2002. Mismatch negativity results from bilateral asymmetric dipole sources in the frontal and temporal lobes. Brain Topogr 15, 13-27.

Jirsa, V.K., 2004. Connectivity and dynamics of neural information processing. Neuroinformatics 2, 183-204.

Jirsa, V.K., Haken, H., 1996. Field Theory of Electromagnetic Brain Activity. Phys Rev Lett 77, 960-963.

Jirsa, V.K., Haken, H., 1997. A derivation of a macroscopic field theory of the brain from the quasi-microscopic neural dynamics. Physica D 99, 503-526.

Kass, R., Raftery, A., 1995. Bayes factors. Journal of the American Statistical Association, 773-795.

Kepecs, A., Uchida, N., Mainen, Z.F., 2006. The sniff as a unit of olfactory processing. Chem Senses 31, 167-179.

Kiebel, S.J., David, O., Friston, K.J., 2006. Dynamic causal modelling of evoked responses in EEG/MEG with lead field parameterization. NeuroImage 30, 1273-1284.

Kiebel, S.J., Garrido, M.I., Friston, K.J., 2007a. Dynamic causal modelling of evoked responses: the role of intrinsic connections. NeuroImage 36, 332-345.

Kiebel, S.J., Garrido, M.I., Moran, R.J., Friston, K.J., 2008. Dynamic causal modelling for EEG and MEG. Cogn Neurodyn 2, 121-136.

Kiebel, S.J., Kloppel, S., Weiskopf, N., Friston, K.J., 2007b. Dynamic causal modelling: a generative model of slice timing in fMRI. NeuroImage 34, 1487-1496.

Kilner, J.M., Mattout, J., Henson, R., Friston, K.J., 2005. Hemodynamic correlates of EEG: a heuristic. Neuroimage 28, 280-286.

Kincses, W.E., Braun, C., Kaiser, S., Elbert, T., 1999. Modelling extended sources of event-related potentials using anatomical and physiological constraints. Hum Brain Mapp 8, 182-193.

Kisvárday, Z.F., Bonhoeffer, T., Kim, D.-S., Eysel, U. , 1996. Functional topography of horizontal neural networks in cat visual cortex (area 18). In: Aertesen, A., Braitenberg, B. (Ed.), Brain theory- biological and computational theories. Elsevier, Amsterdam, pp. 97-122.

Knight, B.W., 1972a. Dynamics of encoding in a population of neurons. J Gen Physiol 59, 734-766.

Knight, B.W., 1972b. The relationship between the firing rate of a single neuron and the level of activity in a population of neurons. Experimental evidence for resonant enhancement in the population response. J Gen Physiol 59, 767-778.

Knight, B.W., 2000. Dynamics of encoding in neuron populations: some general mathematical features. Neural Comput 12, 473-518.

Kopell, N., Ermentrout, G.B., Whittington, M.A., Traub, R.D., 2000. Gamma rhythms and beta rhythms have different synchronization properties. Proc Natl Acad Sci U S A 97, 1867-1872.

Kullback, S., Leibler, R., 1951. On information and sufficiency. The Annals of Mathematical Statistics, 79-86.

Kumar, S., Stephan, K.E., Warren, J.D., Friston, K.J., Griffiths, T.D., 2007. Hierarchical processing of auditory objects in humans. PLoS Comput Biol 3, e100.

Laird, N.M., Ware, J.H., 1982. Random-effects models for longitudinal data. Biometrics 38, 963-974.

Leff, A.P., Schofield, T.M., Stephan, K.E., Crinion, J.T., Friston, K.J., Price, C.J., 2008. The cortical dynamics of intelligible speech. J Neurosci 28, 13209-13215.

Lerner, Y., Hendler, T., Malach, R., 2002. Object-completion effects in the human lateral occipital complex. Cereb Cortex 12, 163-177.

Liley, D.T., Bojak, I., 2005. Understanding the transition to seizure by modelling the epileptiform activity of general anesthetic agents. J Clin Neurophysiol 22, 300-313.

Linden, D.E., Bittner, R.A., Muckli, L., Waltz, J.A., Kriegeskorte, N., Goebel, R., Singer, W., Munk, M.H., 2003. Cortical capacity constraints for visual working memory: dissociation of fMRI load effects in a fronto-parietal network. NeuroImage 20, 1518-1530.

Lisman, J.E., Idiart, M.A., 1995. Storage of 7 +/- 2 short-term memories in oscillatory subcycles. Science 267, 1512-1515.

Litvak, V., Zeller, D., Oostenveld, R., Maris, E., Cohen, A., Schramm, A., Gentner, R., Zaaroor, M., Pratt, H., Classen, J., 2007. LTP-like changes induced by paired associative stimulation of the primary somatosensory cortex in humans: source analysis and associated changes in behaviour. Eur J Neurosci 25, 2862-2874.

Lopes da Silva, F.H., Hoeks, A., Smits, H., Zetterberg, L.H., 1974. Model of brain rhythmic activity. The alpha-rhythm of the thalamus. Kybernetik 15, 27-37.

Lopes da Silva, F.H., van Rotterdam, A., Barts, P., van Heusden, E., Burr, W., 1976. Models of neuronal populations: the basic mechanisms of rhythmicity. Prog Brain Res 45, 281-308.

MacKay, D., 2003. Information theory, inference and learning algorithms. Cambridge University Press.

Maguire, E.A., Gadian, D.G., Johnsrude, I.S., Good, C.D., Ashburner, J., Frackowiak, R.S., Frith, C.D., 2000. Navigation-related structural change in the hippocampi of taxi drivers. Proc Natl Acad Sci U S A 97, 4398-4403.

Mainen, Z.F., Sejnowski, T.J., 1995. Reliability of spike timing in neocortical neurons. Science 268, 1503-1506.

Makeig, S., Westerfield, M., Jung, T.P., Enghoff, S., Townsend, J., Courchesne, E., Sejnowski, T.J., 2002. Dynamic brain sources of visual evoked responses. Science 295, 690-694.

Malenka, R.C., Bear, M.F., 2004. LTP and LTD: an embarrassment of riches. Neuron 44, 5-21.

Manwani, A., and Koch, C., 1999. Detecting and estimating signals in noisy cable structures: I. Neuronal noise sources. Neural Comput., 11, 1797–1829.

Marreiros, A.C., Daunizeau, J., Kiebel, S.J., Friston, K.J., 2008a. Population dynamics: variance and the sigmoid activation function. Neuroimage 42, 147-157.

Marreiros, A.C., Kiebel, S.J., Daunizeau, J., Harrison, L.M., Friston, K.J., 2009. Population dynamics under the Laplace assumption. Neuroimage 44, 701-714.

Marreiros, A.C., Kiebel, S.J., Friston, K.J., 2008b. Dynamic causal modelling for fMRI: a two-state model. NeuroImage 39, 269-278.

Martin, S.J., Grimwood, P.D., Morris, R.G., 2000. Synaptic plasticity and memory: an evaluation of the hypothesis. Annu Rev Neurosci 23, 649-711.

Massimini, M., Ferrarelli, F., Huber, R., Esser, S.K., Singh, H., Tononi, G., 2005. Breakdown of cortical effective connectivity during sleep. Science 309, 2228-2232.

Matthews, P.M., Honey, G.D., Bullmore, E.T., 2006. Applications of fMRI in translational medicine and clinical practice. Nat Rev Neurosci 7, 732-744.

Mattia, M., Del Giudice, P., 2004. Finite-size dynamics of inhibitory and excitatory interacting spiking neurons. Phys Rev E Stat Nonlin Soft Matter Phys 70, 052903.

McIntosh, A.R., 2000. Towards a network theory of cognition. Neural Netw 13, 861-870.

McIntosh, A.R., Grady, C.L., Ungerleider, L.G., Haxby, J.V., Rapoport, S.I., Horwitz, B., 1994. Network analysis of cortical visual pathways mapped with PET. J Neurosci 14, 655-666.

McKeown, M., Jung, T., Makeig, S., Brown, G., Kindermann, S., Lee, T., Sejnowski, T., 1998. Spatially independent activity patterns in functional MRI data during the Stroop color-naming task. National Acad Sciences, pp. 803-810.

McIntosh, A., Gonzalez-Lima, F., 1994. Structural equation modelling and its application to network analysis in functional brain imaging. Human Brain Mapping 2, 2-22.

Mechelli, A., Crinion, J.T., Long, S., Friston, K.J., Lambon Ralph, M.A., Patterson, K., McClelland, J.L., Price, C.J., 2005. Dissociating reading processes on the basis of neuronal interactions. J Cogn Neurosci 17, 1753-1765.

Mechelli, A., Price, C.J., Friston, K.J., Ishai, A., 2004. Where bottom-up meets top-down: neuronal interactions during perception and imagery. Cereb Cortex 14, 1256-1265.

Mechelli, A., Price, C.J., Noppeney, U., Friston, K.J., 2003. A dynamic causal modelling study on category effects: bottom-up or top-down mediation? J Cogn Neurosci 15, 925-934.

Miller, P., Wang, X.J., 2006. Power-law neuronal fluctuations in a recurrent network model of parametric working memory. J Neurophysiol 95, 1099-1114.

Moran, R.J., Kiebel, S.J., Stephan, K.E., Reilly, R.B., Daunizeau, J., Friston, K.J., 2007. A neural mass model of spectral responses in electrophysiology. NeuroImage 37, 706-720.

Moran, R.J., Stephan, K.E., Kiebel, S.J., Rombach, N., O'Connor, W.T., Murphy, K.J., Reilly, R.B., Friston, K.J., 2008. Bayesian estimation of synaptic physiology from the spectral responses of neural masses. NeuroImage 42, 272-284.

Moran, R.J., Stephan, K.E., Seidenbecher, T., Pape, H.C., Dolan, R.J., Friston, K.J., 2009. Dynamic causal models of steady-state responses. NeuroImage 44, 796-811.

Mormann, F., Fell, J., Axmacher, N., Weber, B., Lehnertz, K., Elger, C.E., Fernandez, G., 2005. Phase/amplitude reset and theta-gamma interaction in the human medial temporal lobe during a continuous word recognition memory task. Hippocampus 15, 890-900.

Morris, C., Lecar, H., 1981. Voltage oscillations in the barnacle giant muscle-fibre. Biophysical Journal 35, 193-213.

Mourao-Miranda, J., Bokde, A.L., Born, C., Hampel, H., Stetter, M., 2005. Classifying brain states and determining the discriminating activation patterns: Support Vector Machine on functional MRI data. Neuroimage 28, 980-995.

Naatanen, R., Tervaniemi, M., Sussman, E., Paavilainen, P., Winkler, I., 2001. "Primitive intelligence" in the auditory cortex. Trends Neurosci 24, 283-288.

Neal, R.M., Hinton, G.E., 1998. A view of the EM algorithm that justifies incremental, sparse, and other variants. Learning in Graphical Models 89, 355-368.

Noppeney, U., Price, C.J., Penny, W.D., Friston, K.J., 2006. Two distinct neural mechanisms for category-selective responses. Cereb Cortex 16, 437-445.

Nunez, P., 1974. Brain wave-equation - Model for EEG. Electroencephalography and Clinical Neurophysiology 37, 426-426.

Nykamp, D.Q., Tranchina, D., 2000. A population density approach that facilitates large-scale modelling of neural networks: analysis and an application to orientation tuning. J Comput Neurosci 8, 19-50.

Omurtag, A., Knight, B.W., Sirovich, L., 2000. On the simulation of large populations of neurons. J Comput Neurosci 8, 51-63.

Opitz, B., Rinne, T., Mecklinger, A., von Cramon, D.Y., Schroger, E., 2002. Differential contribution of frontal and temporal cortices to auditory change detection: fMRI and ERP results. NeuroImage 15, 167-174.

Palva, J.M., Palva, S., Kaila, K., 2005. Phase synchrony among neuronal oscillations in the human cortex. J Neurosci 25, 3962-3972.

Penny, W.D., Duzel, E., Miller, K.J., Ojemann, J.G., 2008. Testing for nested oscillation. J Neurosci Methods 174, 50-61.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004a. Comparing dynamic causal models. NeuroImage 22, 1157-1172.

Penny, W.D., Stephan, K.E., Mechelli, A., Friston, K.J., 2004b. Modelling functional integration: a comparison of structural equation and dynamic causal models. NeuroImage 23 Suppl 1, S264-274.

Phillips, C., Zeki, S., Barlow, H., 1984. Localization of function in the cerebral cortex: past, present and future. Brain 107, 328.

Pitt, M.A., Myung, I.J., 2002. When a good fit can be bad. Trends in Cognitive Sciences 6, PII S1364-6613(1302)01964-01962.

Posner, M.I., Sheese, B.E., Odludas, Y., Tang, Y., 2006. Analyzing and shaping human attentional networks. Neural Netw 19, 1422-1429.

Press, W.H., Flannery, B. P., Teukolsky, S. A., Vetterling, W. T., 1999. Numerical Recipes in C: The Art of Scientific Computing. Cambridge University Press.

Raftery, A.E., 1995. Bayesian model selection in social research. Sociological Methodology 1995, Vol 25 25, 111-163.

Ravazzani, P., Tognola, G., Grandori, F., Budai, R., Locatelli, T., Cursi, M., Di Benedetto, G., Comi, G., 1995. Temporal segmentation and multiple-source analysis of short-latency median nerve SEPs. J Med Eng Technol 19, 70-76.

Rinne, T., Alho, K., Ilmoniemi, R.J., Virtanen, J., Näätänen, R., 2000. Separate time behaviors of the temporal and frontal mismatch negativity sources. NeuroImage 12, 14-19.

Risken, H., 1996. The Fokker-Planck equation: Methods of Solutions and Applications Springer, Berlin.

Robinson, P.A., 2005. Propagator theory of brain dynamics. Phys Rev E Stat Nonlin Soft Matter Phys 72, 011904.

Robinson, P.A., Rennie, C.J., Rowe, D.L., O'Connor, S.C., Gordon, E., 2005. Multiscale brain modelling. Philos Trans R Soc Lond B Biol Sci 360, 1043-1050.

Robinson, P.A., Rennie, C.J., Rowe, D.L., O'Connor, S.C., Wright, J.J., Gordon, E., Whitehouse, R.W., 2003. Neurophysical modelling of brain dynamics. Neuropsychopharmacology 28 Suppl 1, S74-79.

Robinson, P.A., Rennie, C.J., Wright, J.J., 1997. Propagation and stability of waves of electrical activity in the cerebral cortex. Physical Review E 56, 826-840.

Robinson, P.A., Rennie, C.J., Wright, J.J., Bahramali, H., Gordon, E., Rowe, D.L., 2001. Prediction of electroencephalographic spectra from neurophysiology. Phys Rev E Stat Nonlin Soft Matter Phys 63, 021903.

Rodrigues, S., Terry, J.R., Breakspear, M., 2006. On the genesis of spikewave activity in a mean-field model of human thalamic and corticothalamic dynamics. Phys. Lett. A 4-5, 352-357.

Rodriguez, R., Tuckwell, H.C., 1998. Noisy spiking neurons and networks: useful approximations for firing probabilities and global behavior. Biosystems 48, 187-194.

Rodriguez, R., Tuckwell, H.C., 2000. A dynamical system for the approximate moments of nonlinear stochastic models of spiking neurons and networks. Mathematical and Computer Modelling 31, 175-180.

Rodriguez, R., Tuckwell, H. C., 1996. Statistical properties of stochastic nonlinear dynamical models of single spiking neurons and neural networks. Phys. Rev. E 54, 5585 - 5590.

Rodriguez, R., Tuckwell, H. C., 1998. A Dynamical System for the Approximate Moments of Nonlinear Stochastic Models of Spiking Neurons and Networks. Mathematical and Computer Modelling 31, 175-180.

Roebroeck, A., Formisano, E., Goebel, R., 2005. Mapping directed influence over the brain using Granger causality and fMRI. NeuroImage 25, 230-242.

Rushworth, M.F., Behrens, T.E., Johansen-Berg, H., 2006. Connection patterns distinguish 3 regions of human parietal cortex. Cereb Cortex 16, 1418-1430.

Shulman, R.G., Hyder, F., Rothman, D.L., 2002. Biophysical basis of brain activity: implications for neuroimaging. Q Rev Biophys 35, 287-325.

Sirovich, L., 2003. Dynamics of neuronal populations: eigenfunction theory; some solvable cases. Network 14, 249-272.

Smith, A.P., Stephan, K.E., Rugg, M.D., Dolan, R.J., 2006. Task and content modulate amygdala-hippocampal connectivity in emotional retrieval. Neuron 49, 631-638.

Sompolinsky, H., Zippelius, A., 1982. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. Physical Review B 25, 6860-6875.

Sotero, R.C., Trujillo-Barreto, N.J., Iturria-Medina, Y., Carbonell, F., Jimenez, J.C., 2007. Realistically coupled neural mass models can generate EEG rhythms. Neural Comput 19, 478-512.

Sporns, O., Tononi, G., Kotter, R., 2005. The human connectome: A structural description of the human brain. PLoS Comput Biol 1, e42.

Stephan, K.E., 2004. On the role of general system theory for functional neuroimaging. J Anat 205, 443-470.

Stephan, K.E., Baldeweg, T., Friston, K.J., 2006. Synaptic plasticity and dysconnection in schizophrenia. Biol Psychiatry 59, 929-939.

Stephan, K.E., Harrison, L.M., Kiebel, S.J., David, O., Penny, W.D., Friston, K.J., 2007a. Dynamic causal models of neural system dynamics:current state and future extensions. J Biosci 32, 129-144.

Stephan, K.E., Kasper, L., Harrison, L.M., Daunizeau, J., den Ouden, H.E., Breakspear, M., Friston, K.J., 2008. Nonlinear dynamic causal models for fMRI. NeuroImage 42, 649-662.

Stephan, K.E., Marshall, J.C., Penny, W.D., Friston, K.J., Fink, G.R., 2007b. Interhemispheric integration of visual processing during task-driven lateralization. J Neurosci 27, 3512-3522.

Stephan, K.E., Penny, W.D., Daunizeau, J., Moran, R.J., Friston, K.J., 2009. Bayesian model selection for group studies. NeuroImage.

Stephan, K.E., Penny, W.D., Marshall, J.C., Fink, G.R., Friston, K.J., 2005. Investigating the functional role of callosal connections with dynamic causal models. Ann N Y Acad Sci 1064, 16-36.

Stephan, K.E., Weiskopf, N., Drysdale, P.M., Robinson, P.A., Friston, K.J., 2007c. Comparing hemodynamic models with DCM. NeuroImage 38, 387-401.

Steriade, M., 2006. Grouping of brain rhythms in corticothalamic systems. Neuroscience 137, 1087-1106.

Steyn-Ross, M., Steyn-Ross, D., Sleigh, J., Liley, D., 1999. Theoretical electroencephalogram stationary spectrum for a white-noise-driven cortex: Evidence for a general anesthetic-induced phase transition. PHYSICAL REVIEW-SERIES E-60, 7299-7311.

Summerfield, C., Egner, T., Greene, M., Koechlin, E., Mangels, J., Hirsch, J., 2006. Predictive codes for forthcoming perception in the frontal cortex. Science 314, 1311-1314.

Summerfield, C., Koechlin, E., 2008. A neural representation of prior information during perceptual inference. Neuron 59, 336-347.

Swanson, L.W., 2003. Brain Architecture. Oxford University Press.

Sychra, J., Bandettini, P., Bhattacharya, N., Lin, Q., 1994. Synthetic images by subspace transforms I. Principal components images and related filters. MEDICAL PHYSICS-LANCASTER PA- 21, 193-193.

Tallon-Baudry, C., Bertrand, O., 1999. Oscillatory gamma activity in humans and its role in object representation. Trends Cogn Sci 3, 151-162.

Tass, P., Rosenblum, M.G., Weule, J., Kurths, J., Pikovsky, A., Volkmann, J., Schnitzler, A., Freund, H.J., 1998. Detection of n : m phase locking from noisy data: Application to magnetoencephalography. Physical Review Letters 81, 3291-3294.

Tass, P.A., 2003. Stochastic phase resetting of stimulus-locked responses of two coupled oscillators: transient response clustering, synchronization, and desynchronization. Chaos 13, 364-376.

Todd, J.J., Marois, R., 2004. Capacity limit of visual short-term memory in human posterior parietal cortex. Nature 428, 751-754.

Trachtenberg, J.T., Chen, B.E., Knott, G.W., Feng, G., Sanes, J.R., Welker, E., Svoboda, K., 2002. Long-term in vivo imaging of experience-dependent synaptic plasticity in adult cortex. Nature 420, 788-794.

Tuckwell, H.C., Rodriguez, R., 1998. Analytical and simulation results for stochastic Fitzhugh-Nagumo neurons and neural networks. J Comput Neurosci 5, 91-113.

Ungerleider, L.G., Haxby, J.V., 1994. 'What' and 'where' in the human brain. Curr Opin Neurobiol 4, 157-165.

Valdes, P.A., Jimenez, J.C., Riera, J., Biscay, R., Ozaki, T., 1999. Nonlinear EEG analysis based on a neural mass model. Biol Cybern 81, 415-424.

Vanhatalo, S., Palva, J.M., Holmes, M.D., Miller, J.W., Voipio, J., Kaila, K., 2004. Infraslow oscillations modulate excitability and interictal epileptic activity in the human cortex during sleep. Proc Natl Acad Sci U S A 101, 5053-5057.

Varela, F., Lachaux, J.P., Rodriguez, E., Martinerie, J., 2001. The brainweb: Phase synchronization and large-scale integration. Nature Reviews Neuroscience 2, 229-239.

von Stein, A., Chiang, C., Konig, P., 2000. Top-down processing mediated by interareal synchronization. Proc Natl Acad Sci U S A 97, 14748-14753.

Wen, Q., Chklovskii, D.B., 2005. Segregation of the brain into gray and white matter: a design minimizing conduction delays. PLoS Comput Biol 1, e78.

Wendling, F., Bellanger, J.J., Bartolomei, F., Chauvel, P., 2000. Relevance of nonlinear lumped-parameter models in the analysis of depth-EEG epileptic signals. Biol Cybern 83, 367-378.

Wilson, H.R., Cowan, J.D., 1972. Excitatory and inhibitory interactions in localized populations of model neurons. Biophys J 12, 1-24.

Wilson, H.R., Cowan, J.D., 1973. A mathematical theory of the functional dynamics of cortical and thalamic nervous tissue. Kybernetik 13, 55-80.

Wolters, A., Schmidt, A., Schramm, A., Zeller, D., Naumann, M., Kunesch, E., Benecke, R., Reiners, K., Classen, J., 2005. Timing-dependent plasticity in human primary somatosensory cortex. J Physiol 565, 1039-1052.

Worsley, K., Liao, C., Aston, J., Petre, V., Duncan, G., Morales, F., Evans, A., 2002. A general statistical analysis for fMRI data. NeuroImage 15, 1-15.

Worsley, K.J., Marrett, S., Neelin, P., Vandal, A.C., Friston, K.J., Evans, A.C., 1996. A unified statistical approach for determining significant signals in images of cerebral activation. Human Brain Mapp 4, 58-73.

Wright, J.J., Liley, D.T.J., 1996. Dynamics of the brain at global and microscopic scales: Neural networks and the EEG. Behavioural and Brain Sciences 19, 285-&.

Wright, J.J., Rennie, C.J., Lees, G.J., Robinson, P.A., Bourke, P.D., Chapman, C.L., Gordon, E., Rowe, D.L., 2003. Simulated electrocortical activity at microscopic, mesoscopic, and global scales. Neuropsychopharmacology 28 Suppl 1, S80-93.

Zetterberg, L.H., Kristiansson, L., Mossberg, K., 1978. Performance of a model for a local neuron population. Biol Cybern 31, 15-26.